AD_____

Award Number: DAMD17-96-1-6042

TITLE: Dysregulation of the Stress Response in the Persian Gulf Syndrome

PRINCIPAL INVESTIGATOR: Daniel Clauw, M.D.

CONTRACTING ORGANIZATION: Georgetown University Medical Center
Washington, DC 20007

REPORT DATE: November 1999

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

-REPORT DOCUMENTATION PAGE-

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1999	Final (6 May 96 – 5 Oct 99)

4. TITLE AND SUBTITLE
Dysregulation of the Stress Response in the Persian Gulf Syndrome

5. FUNDING NUMBERS
DAMD17- 96-1-6042

6. AUTHOR(S)
Daniel Clauw, M.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Georgetown University Medical Center
Washington, DC 20007

e-mail:
clauwd@gunet.georgetown.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distributed unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

Approximately 20% of Gulf War veterans who have presented to DoD and VA health registries have unexplained symptom-based illnesses that have been termed the Persian Gulf Syndrome or Gulf War Illnesses (GWI). Similar syndromes (fibromyalgia (FMS), chronic fatigue syndrome (CFS), etc.) are also known to occur at a high rate in the general population. Collectively, we refer to these syndromes as chronic multisymptom illnesses (CMI). CMI are typically initiated and perpetuated by a variety of physical and emotional stressors. Studies of CMI have shown that there are a number of objective neurohumoral abnormalities in the human "stress response" which may be responsible for the symptoms seen in these entities. This study was designed to demonstrate that individuals with GWI: 1) display centrally mediated disturbances in autonomic tone, leading to smooth muscle dysmotility, and symptoms such as irritable bowel syndrome, 2) display diffuse disturbances in nociception (pain threshold) that are partly responsible for many of the pain-related symptoms seen in GWI, and 3) display the same blunting of the hypothalamic-pituitary axes seen in some CMI, and contributes to the observed fatigue. We have extensively studied these three different stress responses in a total of 125 subjects in this project (25 GWI, 49 healthy normal controls, and 51 with CMI). Currently our data demonstrate that this cohort of GWI subjects shows evidence of peripheral nociceptive abnormalities, as well as smooth muscle dysmotility (similar process that may underly CMI).

14. SUBJECT TERMS
Gulf War Illness

15. NUMBER OF PAGES
53

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

_____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

_____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

✓ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

10/25/00

PI - Signature Date

Table of Contents

FINAL REPORT (REVISED)

April 23, 2001

Award Number DAMD17-96-1-6042
Dysregulation of the Stress Response in the Persian Gulf Syndrome

Daniel J. Clauw, M.D.
Principal Investigator

INTRODUCTION (adapted from original application)

In 1990 and 1991, the U.S. deployed approximately 700,000 troops to the Persian Gulf to liberate Kuwait from Iraqi occupation. Fortunately, there were relatively few combat and non-combat related injuries and diseases during this conflict in comparison with previous military campaigns, and most veterans of this conflict who did develop illness had diagnosable and treatable conditions [1,2]. However, the illnesses associated with approximately 20% of those with symptoms have not been explained, and this constellation of symptoms occurring in this setting has been termed the Persian Gulf Syndrome or Gulf War Illnesses (GWI).

We have hypothesized that the only unique aspect of GWI was the location and timing of troop deployment. Similar illnesses have been noted after nearly every major conflict, although these syndromes have had different names and attributions. *More importantly, similar syndromes occur at a high rate in the general population, with the currently preferred terms being fibromyalgia (FMS), chronic fatigue syndrome (CFS), somatoform disorder, and multiple chemical sensitivity (MCS).*

We have been extensively involved in the study of these latter illnesses, and there are substantial data suggesting that these syndromes are not discrete entities, but rather fall within a continuum. All of these illnesses are typically initiated and perpetuated by a variety of physical and emotional stressors, as may have occurred in deployment to the Gulf War and upon return home. The study of these illnesses has shown that there are a number of objective abnormalities in the human stress response that can be identified that may be responsible for the symptoms seen in these entities.

The purpose of this proposal was to demonstrate that these same objective neurohormonal abnormalities are present in individuals with GWI, and may be at least partly responsible for the symptoms noted in these individuals. There are several axes of the stress response that can independently or concurrently function aberrantly in these conditions, including the autonomic nervous system, the hypothalamic-pituitary axes, and descending anti-nociceptive pathways. Specific symptomology results when each of these systems function improperly. In this study we propose to demonstrate that: 1) individuals with GWI display centrally mediated disturbances in autonomic tone, and this leads to vasomotor instability and smooth muscle dysmotility, and symptoms such as irritable bowel syndrome, and migraine headaches, 2) individuals with GWI display diffuse disturbances in nociception (pain threshold) that are partly responsible for many of the pain-related symptoms seen in this condition (e.g., myalgia, arthralgia, sore throat), and 3) the same neuroendocrine changes seen in FMS/CFS and CFS, are seen in GWI, and contribute substantially to the fatigue seen in this condition.

We feel that this study may lead to important insights into the symptoms experienced by some Persian Gulf veterans. The demonstration of common underlying pathophysiologic mechanisms in FMS, CFS, and GWI will be tremendously beneficial, in that this should lead both to more effective treatment of these individuals, and perhaps to effective strategies regarding avoidance of this problem in future conflicts.

EXPERIMENTAL METHODS (adapted from original application)

Overview. The current study was conducted on individuals who were admitted to the Washington, D.C. VAMC, one of the three VA Persian Gulf Referral Centers created in August of 1992. This project was a multidisciplinary collaborative effort, involving individuals from Georgetown University Medical Center and the VAMC, as well as consultants from the National Institutes of Health who are recognized as international experts on the effects of stress on the neuroendocrine and autonomic systems. The subjects in this study were admitted to the Clinical Research Center (CRC) at Georgetown for two days. Over the course of two days, participants underwent a series of studies that permit the concurrent evaluation of a number of physiologic and biochemical parameters.

The physiologic studies performed measure both the qualitative and quantitative aspects of a number of symptoms, and include specialized testing of peripheral and visceral nociception, and smooth muscle motility. We also evaluate multiple indices of autonomic function, including neurohormone levels at baseline and after standardized stressors.

We are employing several Control groups: 1) "healthy normals", 2) Gulf War Illness (GWI) patients without a symptom or feature being studied, and 3) individuals with CFS and FMS. The purpose of the healthy normal Control group was to show that the GWI patients differ from age and gender matched Controls; this is the most common type of Control group employed for this type of study. We will also compare the results of the pathophysiologic studies with the same tests performed concurrently in FMS/CFS and CFS, to demonstrate that there are no differences between the GWI patients and the FMS/CFS groups.

Subject recruitment. We originally planned to evaluate 40 consecutive GWI veterans who were referred to the Washington VAMC for a Referral Examination, and 20 age- and gender-matched healthy Controls, and 20 persons who have diagnosed FMS/CFS. Studying individuals who were being admitted to the VAMC for a Referral Examination had several theoretical advantages over the use of alternative sources of GWI patients: 1) individuals in our study would be well screened for alternative causes of symptoms, and those individuals with an alternative illness would be precluded from participation, and 2) these individuals would have extensive baseline testing performed as part of the Referral Examination which would be available for analysis.

a) *Definition of GWI.* Because there was no widely-accepted definition of GWI, this was a difficult issue. We sought to develop a definition of GWI that would truly select individuals who served in the Gulf War who have a significant unexplained illness. The definition we employed for purposes of this study was that: 1) unexplained symptoms developed within 6 months of participation in the Gulf War, and continue to be active, and 2) those symptoms include 3 or more of the following: myalgia, arthralgia, headache, fatigue severe enough to limit activities, cognitive dysfunction, sensitivity to multiple environmental substances, pulmonary symptoms, and GI symptoms.

b) *Inclusion and exclusion criteria.* Entry and exclusion criteria, other than meeting the above noted GWI criteria, included: *1)* ages 18 to 60, and *2)* subjects must not consume any antidepressant, tricyclic compound, benzodiazepine, anti-inflammatory, or antipsychotic medication for two weeks prior to study (these drugs interfere with testing being performed).

Control recruitment. Twenty sedentary healthy individuals, and twenty patients with diagnosed FMS and/or CFS, who were matched for age and gender to represent the study population were randomly selected. These individuals were compensated for participation, and the primary source of Controls was employees and patients at Georgetown University Medical Center.

Methods. Eligible subjects were given informed consent and scheduled for admission to the Georgetown University Medical Center Clinical Research Center (CRC). Testing throughout the day occurred in the CRC, except for the gastroenterology studies. The schedule of testing is listed below in Table 1, and the sequence was identical for subjects and Controls. Methods for each test, and justification where appropriate, are described in detail below.

Table 1 – Schedule of testing

	Day One	Day Two
MORNING	Serum and blood collection (8AM) [Begin 24-hour urine collection and Holter monitoring] Tender point examination Tilt table testing COGSCREEN	Serum and blood collection (8AM) [Remove Holter monitor and complete 24 hour urine collection] Dolorimeter exam Gastroenterology evaluation
AFTERNOON	Structured Clinical Interview (SCID) Serum collection (4PM)	
EVENING	Complete self-report questionnaires	

Serum collection. Serum and plasma was collected utilizing standard venipuncture techniques. As noted, blood was drawn at standardized times throughout the study to eliminate discrepancies due to diurnal variation. Samples were placed on ice immediately and kept dark until they were centrifuged. Sera were distributed into several aliquots and stored at -70°C at two different locations.

Dolorimeter examination. A dolorimeter is a simple mechanical pressure gauge designed to quantify pain threshold and tolerance. For this examination, the pressure (in kg) necessary to produce discomfort (pain threshold) and unbearable pain (pain tolerance) at 18 designated tender points and 4 control points was recorded. This testing produced five variables for each subject: tender point threshold and tolerance, control point threshold and tolerance, and tender point

count. Since there was such a high correlation between the first four values (r>.85 in our pilot studies), we utilized the tender point threshold as the measure of peripheral pain for primary data analysis. This test had the least variance and best ability to separate FMS/CFS patients from Controls in our pilot data.

Autonomic function.

Overview The assessment of autonomic function in humans is complex. There are no tests of either sympathetic or parasympathetic function that are all-encompassing. In ordinary situations, there are a number of factors controlling visceral motor function, including central autonomic input (which we are measuring), nonadrenergic noncholinergic (NANC) nerves, local reflex loops, and local neurochemical effects. Because symptoms suggestive of smooth muscle dysmotility occur in several organs in CFS, we hypothesize that *aberrant centrally mediated autonomic input* is the predominant stimulus for the abnormal motility. Although the tests we have chosen will assess total autonomic tone we will focus on those tests which measure the central component of autonomic tone.

In addition to testing basal central autonomic tone, the studies chosen also tested the response of the autonomic nervous system to standardized physiologic stressors (mental concentration during cognitive testing, tilt table testing). We tested neuroendocrine function and smooth muscle motility in the same manner. We feel that this was a significant strength of the study, because all clinical and laboratory evidence in FMS/CFS suggests an inability to respond normally to physiologic stressors. We proposed that GWI patients will exhibit similar aberrant stress responses. The stressors we have chosen are likely to accentuate the anomalies in the stress response, especially when compared to some techniques such as response to valsalva or deep breathing (autonomic testing), or response to corticotropin releasing hormone (CRH) (neuroendocrine testing).

Holter monitoring. Heart rate variability monitoring has been demonstrated to be an accurate means of assessing both the sympathetic and parasympathetic components of autonomic tone [3]. This can be performed by temporal or spectral analysis the time between successive normal QRS complexes (N-N intervals) over an entire 24 hour period, or over short intervals of time, to determine how the autonomic nervous system functions in response to specific stimuli. Individuals with low efferent parasympathetic tone will display an elevated resting heart rate, and heart rate variation with breathing. Sympathetic tone can likewise be assessed with this technique, with tilt table testing as a useful adjunct in this regard. An especially useful feature of Holter monitoring to assess autonomic tone is that it affords a functional assessment of autonomic function over an extended period of time that includes provocative maneuvers and stressful events (tilt table testing as well as visceral and peripheral nociceptive testing). This was especially important since in the chronic phase of these disorders, we hypothesize that individuals may be particularly impaired in their ability to respond to these stressors.

Patients wore a standard ambulatory ECG recorder for the initial 24-hour period of the study. Data were analyzed using a dedicated Marquette series 8000 analyzer with specialized software for Heart Rate Variability. A diary was kept so that we could analyze how the subjects respond to

each of these stimuli. In addition, this Holter monitor allows the "marking" of "events" (e.g. the onset of tilt table testing, cognitive testing, lumbar puncture) on the tape, so that the autonomic responses to these specific stressors can be analyzed.

Tilt table testing. Tilt table testing allows a standardized method of assessing primarily the sympathetic component to central autonomic tone. With tilting in normals, the systolic blood pressure may fall as much as 15 mm Hg, whereas the diastolic may drop up to 5 mmHg. Any fall of blood pressure greater than this was considered abnormal.

Individuals are supine on a tilt table with foot support and secured in position. The subject rests in this position for at least 10 minutes. Blood pressure and heart rate are recorded using an automatic blood pressure recorder. Blood pressure and heart rate were recorded every minute, and blood samples are drawn every five minutes through minute-15. As with earlier samples, the blood was kept on ice and in the dark until centrifugation. All samples were frozen at -70°C for later batch analysis.

Gastrointestinal evaluation. For evaluation of esophageal smooth muscle tone, and esophageal nociception, we follow the standard protocols for motility studies and Bernstein tests. The results obtained include baseline manometric data (normal or abnormal, based on defined criteria), as well as the results of three provocative tests (chest pain with edrophonium, dysmotility with edrophonium, and modified Bernstein test). Also, the diameter of an esophageal balloon required to elicit pain was recorded. The data utilized in primary data analysis were: 1) the presence or absence of baseline dysmotility, and 2) the diameter of balloon causing nociception, which measures visceral nociception.

Neuroendocrine studies. Samples were collected in a uniform manner so that we could assess the integrity of this system. At the time this grant was planned, most studies had suggested a global blunting of hypothalamic pituitary adrenal function in illnesses such as FMS and CFS. Since that time, the story has become more complex [13], with differences between patient groups and Control groups being uniformly detected, but the direction and magnitude of these differences appears to depend on a number of factors, particularly the setting in which the testing occurs (i.e., at rest, or with exposure to stressors).

When this grant was planned we felt that provocative tests with "releasing" hormones gave a somewhat artificial understanding of neuroendocrine functional status. We felt (and still feel) that a more relevant physiologic test of the neuroendocrine system in a disorder such as FMS/CFS is to determine how subjects respond to standardized physical and emotional, rather than hormonal, stressors. Therefore, we designed this study to look at the level of hormones at given points in time (8AM and 4PM), under standardized testing conditions (including orthostatic and cognitive stressors), to determine both the basal level of these hormones and the capacity to change levels in response to a physiologic stimulus. We collected this same data taken under identical conditions on both GWI patients and Controls, at both baseline and after stressors, which allowed us to assess further the differences between the neuroendocrine function within the group of patients, as well as between GWI patients and Controls.

Structured Clinical Interview (SCID) and psychiatric evaluation. An extensive psychiatric evaluation was performed as part of this study because psychiatric co-morbidity is high in all of the disorders related to GWI, including FMS/CFS, and MCS. As noted previously, psychiatric variables are used in the secondary analysis of data to determine if these are important co-factors in symptom expression or in the outcomes of physiologic studies. The primary purpose of this portion of the evaluation is to 1) determine the presence of *present* psychiatric disorders, 2) determine the presence of *pre-existing* psychiatric diagnoses, and the effect on the expression of symptoms, and 3) measure the intensity of psychiatric variables such as depression, anxiety, somatic amplification, and use these measures as co-variates to dependent variables such as nociception, autonomic function, and neuroendocrine function.

The psychiatric evaluation utilized the following instruments, which have all been extensively validated, and have been utilized in research in FMS/CFS, and allied conditions:

1) A Structured Clinical Interview for DSM-III-R (SCID)[5]. The structured clinical interview was used to generate current and lifetime psychiatric disorders such as mood disorders, anxiety disorders, and somatoform disorders. A PTSD module is included as well. The SCID is widely used as the "gold standard" for psychiatric diagnosis in the research setting. Either Dr. Epstein or a subordinate who has been appropriately trained conducted the interview. Particular attention was paid to the timing of all psychiatric symptoms, including somatization symptoms, as they correspond to service in the Gulf War.

2) Beck Depression Inventory (BDI) [6]. The BDI is a 21-item measure of the severity of current depressive symptoms, including both neurovegetative and cognitive symptoms of depression.

4) The RAND 36 item health survey (SF-36) [7]. This survey is a self-report measure of functional health status that has been widely utilized. Eight domains were assessed: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, social functioning, energy/fatigue, and general health perception.

5) NEO PI-R [8]. The NEO is a self-report inventory for the assessment of personality traits based on a five factor model of personality.

6) Barsky Amplification Scale [9]. This ten item scale measures somatosensory amplification, the tendency to experience somatic sensations as unusually intense or disturbing.

7) Multidimensional Fatigue Inventory . This is a 20-item self-report scale measuring different aspects of fatigue.

8) Short-form McGill Pain Questionnaire [11]. Consists of 15 descriptors providing information on the sensory, affective, and evaluative dimensions of pain experience and is capable of discriminating among different pain problems.

Self-report questionnaires. Subjects completed a packet of self-report questionnaires designed to evaluate assorted symptoms and perceptions regarding pain and fatigue.

Urine collection. Urine was collected during the initial 24-hour period of the study, for archival storage in potential determination of neurohormonal levels, as noted above.

Computerized Cognitive Testing. This COGSREEN battery takes approximately 50 minutes to administer, and has been validated as a sensitive measure of cognition, especially in the areas of interest (attention and short-term memory) [12]. In this study, we used computerized cognitive testing as a psychological stressor, not to test any hypotheses regarding cognition in FMS/CFS. There are considerable animal data suggesting that the physiologic response to physical and emotional stressors may be quite different, so we have incorporated both into our study design.

Sample Size Calculations. A sample size of 40 GWI, 20 FMS/CFS patients, and 20 healthy normal Controls was chosen for this study, and was more than adequate to test the primary hypotheses. For testing differences among means we were seeking to identify a difference of one or more standard deviations between groups as statistically significant (p=.05). Differences of less than one standard deviation in this context were considered unlikely to be biologically meaningful. The standard computation shows 1) that a sample size of 40 patients and 20 Controls gives a power of greater than 90% to detect these differences, even after allowing for moderate adjustments to the p value for multiple hypotheses, and 2) that the projected sample sizes enhanced to include extra cases will produce this same power even when the comparisons are done between two subgroups of the FMS/CFS patients.

Results and Findings:

Recruitment and Subjects

Throughout the study, we had considerable difficulty recruiting GWI subjects, as have other groups of investigators studying this illness. The problems that we encountered which were unique to our site included the fact that the number of persons entering the Washington VAMC Persian Gulf Veterans Referral Center dropped significantly from the time of the grant application, and the staff at the Washington VAMC who agreed to collaborate on this study left the Washington VAMC, leaving us with no collaborators who were committed to this effort.

After encountering these initial difficulties, we made considerable additional efforts to recruit the GWI cohort. These included contacting Gulf War veterans directly instead of counting on our collaborators at the Washington VAMC, offering a free clinical evaluation and treatment recommendations for participants. We also enlisted the help of an additional individual, Dr. David Nashel, who is the Chief of Rheumatology at the Washington VAMC. These efforts brought our total number of GWI subjects to 25, which, with the correspondingly larger comparison cohorts that we recruited, give adequate power to test the hypotheses put forth.

In the three years of the study, a total of 125 subjects were admitted to the Clinical Research Center at Georgetown and run through all or some of the above protocols. Of this total, 25 subjects were individuals who met our criteria for GWI, 49 were healthy normal Controls, and 51 were individuals with chronic multi-symptom illnesses (CMI) such as Fibromyalgia (FMS) and/or Chronic Fatigue Syndrome (CFS). This *n* of 125 is much larger than the total of 80 subjects that we had originally proposed to run through the protocol, although the distribution within the three cohorts is different than what we had originally proposed. Of the 51 CMI

subjects, 13 were taking medications, and although we did not identify differences in dependent or independent variables between the medicated and un-medicated groups, we used only the un-medicated subjects for most analyses.

The three groups were matched for race, and differed only slightly in age (mean ages for GWI 37, and for CMI and Controls, 41). However, the GWI cohort had higher percentages of males than the other two cohorts, and both cohorts of ill individuals demonstrated more psychological co-morbidities. For each data analysis below these differences are dealt with separately, depending on whether gender is found to influence the variable. Table 2 presents age and gender statistics for each group of subjects.

Table 2 – Subject characteristics

	Sex			Age		
	Mean	SD	Median	Mean	SD	Median
Control (N=49)	.12	.33	0	41.8	9.6	45.0
GWI (N=25)	.72	.46	1	36.0	8.6	36.0
FMS/CFS (N = 51)	.06	.24	0	41.1	9.2	44.0

Analytic Strategies

The analytic strategies we employed were selected to provide the clearest comparisons among the study groups while recognizing the limitations imposed by the relative small samples and the confounding effect of some variables (e.g. gender) on our interpretations. It was out of a consideration of these limitations that we did not include all of the possible individual difference variables (e.g. NEO personality scales) in the analyses. Likewise, our decision to form a priori composites in which each component of the composite was given equal weight (for example to assess severity) rather than to undertake multivariate analyses on sets of related variables reflects this consideration. However, performed post-hoc analyses of individual response variables to demonstrate the meaningfulness of using composite scores. Finally, we elected to use the maximum number of subjects available for many analyses rather than restricting ourselves to only those subjects who had complete data. These decisions may have affected the analyses and we address these as they arise.

Questionnaires

Psychiatric Interviews and Questionnaires

SCID

The SCID produces a diagnostic index for lifetime and current diagnoses of 26 mood, anxiety, and personality disorders. The index is on a four point scale with 1=absent, 2=subthreshold, 3=threshold, and 4=not enough information. The SCID was administered to 8 Control, 38 FMS/CFS, and 23 GWI subjects. Using Pearson chi-squared analysis, the results of the SCID showed a general absence of any mood, anxiety, or personality disorders in all three groups of subjects. The one exception is for lifetime major depression. Fifty-one percent of FMS/CFS subjects, 39.1% of GWI subjects, and 0% of Control subjects were diagnosed as being at threshold for lifetime major depression with a marginally significant group effect (p=0.056,

Pearson chi-squared analysis). However, there were no differences among groups for current major depression. Consequently, the meaning of the finding for lifetime major depression in the present context is not clear. Additionally, there were no differences within groups with respect to past or current psychiatric diagnoses

The somatoform module showed a significant group effect (p=0.000) for the numbers of somatoform symptoms. The mean number of somatoform symptom were 10.6±5.7, 9.8±5.9, and 1.6±3.5 for FMS/CFS, GWI, and Control subjects respectively. Both FMS/CFS and GWI subjects had significantly more somatoform symptoms compared with Controls (p=0.000 and p=0.001 respectively. There was no difference in somatoform symptom numbers between FMS/CFS and GWI subjects. The results demonstrate that based on a clinical interview FMS/CFS and GWI subjects both report a greater number of somatic symptoms compared to Control. This finding is borne out below when we examine in detail the self-reported symptoms of subjects.

With respect to current PTSD there were no significant differences among groups. However, eight percent of FMS/CFS subjects and 22% of GWI subjects met criteria for current PTSD compared to 0% of Controls. Although this may represent a real trend toward more PTSD among GWI veterans, the small number of Control subjects makes this difficult to conclude.

NEO PI-R

The NEO PI has five subscales based on a model of personality: *neurotic, extrovert, openness, agreeable, and conscientious.* The NEO-PI was administered to 41 Control, 38 FMS/CFS, and 23 GWI subjects. Table 3 below shows the results for these five subscales in the study groups.

Table 3 - Results of NEO-PI Testing for Personality Traits

	Control		FMS/CFS		GWI		P-values			
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Group**	**Con.-FMS/CFS**	**Con.-GWI**	**FMS/CFS-GWI**
Neurotic	**45.0**	10.3	**51.8**	10.1	**53.4**	13.3	0.004	0.007	0.004	NS
Extrovert	**54.0**	9.7	**46.8**	11.4	**44.8**	12.3	0.002	0.005	0.002	NS
Openness	**57.1**	7.8	**55.8**	11.0	**47.2**	10.7	0.000	NS	0.000	0.001
Agreeable	**53.0**	8.0	**54.2**	11.1	**49.1**	12.2	0.173	NS	NS	0.066
Conscientious	**49.5**	9.6	**49.5**	12.6	**50.4**	13.2	NS	------	-----	-----

Bearing in mind that these data represent only a subset of all questionnaires and evaluations, it is worth noting that both the FMS/CFS and GWI subjects appear to be more neurotic and less extroverted compared to the Control subjects. In addition GWI subjects appear to be less open and agreeable than FMS/CFS subjects.

BDI, STAI, and Barsky

The BDI was administered to 42 Control, 39 FMS/CFS, and 23 GWI subjects; STAI was administered to 41 Control, 40 FMS/CFS, and 24 GWI subjects; Barsky was administered to 42 Control, 39 FMS/CFS, and 22 GWI. Both FMS/CFS and GWI subjects scored significantly higher ($p=0.000$ and $p<0.01$ respectively) than Controls on both the BDI and the state anxiety portion of the STAI. The values for BDI were: 13.7±7.8 for FMS/CFS, 13.4±9.5 for GWI, and 4.1±4.7 for Control. For STAI the values were: 40.6±11.3 for FMS/CFS, 40.8±13.6 for GWI, and 28.5±8.5 for Control. Between FMS/CFS and GWI there were no differences for either BDI or STAI. With all three groups combined, BDI and STAI were highly correlated with one another ($r=0.694$, $p=0.000$). In light of the results from the SCID, these findings suggest that although there were no significant mood, anxiety, or personality disorders diagnosed, the FMS/CFS and GWI subjects both experience about the same level of mood and anxiety symptoms, which are more than observed in Control subjects. On the Barsky amplification scale there were no differences among Controls, FMS, and GWI. However, the SCID shows that both FMS/CFS and GWI subjects report significantly more symptoms in all body systems. The absence of a group effect on Barsky may mean that although FMS/CFS and GWI subjects report more symptoms, the fidelity of that reporting is accurate.

Symptom and General Health Questionnaires

VAS Symptom Questionnaire

An eight-item, self administered questionnaire on somatic symptoms was administered to 46 Control, 44 FMS/CFS, and 25 GWI subjects using a visual analog scale that allowed the subjects to rate symptoms covering a number of body systems on a continuous scale. Table 4 below shows the results of this self-evaluation.

Table 4 – Results of VAS Symptom Questionnaire

	Control		FMS/CFS		GWI		P-values			
	Mean	SD	Mean	SD	Mean	SD	Group	Con.-FMS/CFS	Con.-GWI	FMS/CFS-GWI
Muscle Pain	**0.6**	1.0	**6.4**	2.3	**5.9**	2.3	0.000	0.000	0.000	NS
Muscle Spasm	**0.2**	0.3	**4.2**	3.0	**3.6**	2.9	0.000	0.000	0.000	NS
Muscle Weakness	**0.4**	0.8	**5.0**	3.0	**5.0**	2.3	0.000	0.000	0.000	NS
Numbness & Tingling	**0.2**	0.4	**3.5**	3.2	**4.8**	3.0	0.000	0.000	0.000	NS
Fatigue	**1.4**	1.9	**7.3**	1.9	**7.5**	2.1	0.000	0.000	0.000	NS
Thinking/Memory	**0.7**	1.1	**5.6**	2.7	**6.1**	2.6	0.000	0.000	0.000	NS
Stomach/Intestinal	**0.4**	1.1	**4.8**	3.5	**4.8**	3.7	0.000	0.000	0.000	NS
Breathing Problems	**0.4**	0.9	**2.5**	3.0	**3.3**	2.9	0.000	0.000	0.004	NS

For all symptoms, FMS/CFS and GWI subjects reported significantly higher VAS scores compared with Control subjects. Furthermore, although GWI subjects scored all symptoms higher than FMS/CFS subjects, there were no statistically significant differences observed.

These results reflect the somatic symptom findings of the SCID. Thus, both FMS/CFS and GWI subjects in this study report more symptoms in all body systems to an equal degree

MOS SF-36

Forty–six Control, 44 FMS/CFS, and 24 GWI subjects provided a self-evaluation of their perceived physical and mental health status with the Medical Outcomes Study Short Form 36 (SF-36). The SF-36 measures the following eight health disease concepts:

1. Limitations in physical activities because of health problems (SF-3601);
2. Limitations in usual role activities because of physical health problems (SF-3602);
3. Bodily pain (SF-3603);
4. General health perceptions (SF-3604);
5. Vitality (SF-3605) (energy and fatigue)
6. Limitations in social activities because of physical or emotional problems (SF-3606);
7. Limitations in usual role activities because of emotional problems (SF-3607);
8. Mental health (SF-3608) (psychological distress and well-being).

Table 5 shows the mean, SD, and group statistics for each SF-36 item.

Table 5 – Results of MOS SF-36

	Control		FMS/CFS		GWI		P-values			
	Mean	SD	Mean	SD	Mean	SD	Group	Con.-FMS/CFS	Con-GWI	FMS/CFS-GWI
SF-3601	**94.4**	15.6	**43.4**	28.0	**54.8**	25.1	0.000	0.000	0.000	0.038
SF-3602	**89.7**	18.6	**47.1**	29.6	**49.8**	26.1	0.000	0.000	0.000	0.060
SF-3603	**94.8**	17.6	**12.5**	27.2	**22.9**	27.5	0.000	0.000	0.000	0.007
SF-3604	**86.6**	30.6	**42.7**	39.9	**70.8**	39.7	0.000	0.000	NS	0.012
SF-3605	**81.6**	12.2	**60.0**	18.9	**62.5**	21.5	0.000	0.000	0.001	NS
SF-3606	**67.7**	19.4	**22.3**	18.4	**26.0**	18.4	0.000	0.000	0.000	NS
SF-3607	**86.6**	17.9	**36.9**	21.8	**42.6**	25.1	0.000	0.000	0.000	NS
SF-3608	**83.2**	17.4	**32.4**	17.6	**35.3**	22.4	0.000	0.000	0.000	NS

Both FMS/CFS and GWI have significantly lower scores compared to Controls on all items except for general health perceptions where the GWI scores are not statistically significantly different from Controls. Items 1-4, which focus on physical health GWI scores are significantly higher than FMS/CFS as well as being significantly lower than Controls with their values falling between FMS/CFS and Controls. However, in items 5-8, which focus more on mental health, both FMS/CFS and GWI score equally poorly. The interpretation of this is that GWI subjects perceive their physical health to be better that FMS/CFS subjects do, but mental health perceptions of GWI are equal to those of FMS/CFS subjects

Pain and Fatigue Questionnaires

McGill Pain Questionnaire

Table 6 below shows the results of the self-administered McGill Pain Questionnaire, which consists of 15 descriptors providing information on the sensory, affective, and evaluative dimensions of pain experience and is capable of discriminating among different pain problems. Twenty-six Control, 5 FMS/CFS, and 14 GWI subjects took the McGill Pain Questionnaire.

Table 6 – Results of McGill Pain Questionnaire

	Control		FMS/CFS		GWI		P-values			
	Mean	SD	Mean	SD	Mean	SD	Group	Con.-FMS/CFS	Con-GWI	FMS/CFS-GWI
Pain Presence*	**0.2**	0.4	**1.0**	0.0	**1.0**	0.0	0.000	0.000	0.000	NS
Pain Intensity	**0.5**	0.9	**3.4**	0.9	**2.8**	0.9	0.000	0.000	0.000	NS
VAS Pain Intensity	**0.4**	0.9	**4.3**	1.6	**5.2**	2.1	0.000	0.000	0.000	NS
Pain Total	**1.3**	3.8	**16.4**	11.5	**18.6**	9.2	0.000	0.000	0.000	NS
Sensory Dimension	**1.1**	3.0	**13.2**	8.6	**13.8**	6.9	0.000	0.000	0.000	NS
Affective Dimension	**0.2**	0.9	**3.2**	3.1	**4.7**	2.8	0.000	0.002	0.000	NS

**Pain Presence is a yes(1)/no(0) question. The means for pain presence thus equal the percent of subjects in each group who report pain. Every FMS/CFS and GWI subject reported the presence of pain, compared to only 20% of Controls.*

The McGill Pain Questionnaire shows both FMS/CFS and GWI subjects report significantly more pain than Control subjects and that the ratings in different dimensions of pain are similar between the FMS/CFS and GWI subjects.

Multidimensional Fatigue Inventory (MFI)

The MFI is a 20-item self-report measuring different aspects of fatigue, and was administered to the same subjects as the McGill Pain Questionnaire. Table 7 below shows the results of the MFI.

Table 7 – Results of the Multidimensional Fatigue Inventory

	Control		FMS/CFS		GWI		P-values			
	Mean	SD	Mean	SD	Mean	SD	Group	Con.-FMS/CFS	Con-GWI	FMS/CFS-GWI
General Fatigue	**8.2**	3.8	**16.8**	3.3	**16.0**	3.1	0.000	0.000	0.000	NS
Physical Fatigue	**6.2**	2.2	**15.4**	1.3	**13.2**	4.0	0.000	0.000	0.000	0.042
Reduced Activity	**7.2**	3.0	**12.0**	2.5	**11.6**	4.5	0.000	0.003	0.077	NS
Reduced Motivation	**6.6**	2.3	**11.4**	3.4	**12.0**	4.3	0.000	0.001	0.005	NS
Mental Fatigue	**7.5**	4.3	**12.6**	3.7	**13.6**	4.0	0.002	0.013	0.008	NS

These results show that both FMS/CFS and GWI subjects report greater fatigue in several different dimensions. Although there generally is no difference in reporting between FMS/CFS and GWI, FMS/CFS subjects appear to report greater physical fatigue compared to GWI subjects.

Neuroendocrine Function

Before beginning the analyses we checked for outliers on the dependent variables (see below for a full discussion of outliers). For example, based on significant values of the

Studentized Residual (SR), which can be interpreted as a t value, we eliminated subjects from the ACTH and Cortisol determinations as follows.

ACTH Baseline (8:30AM) Day 1	Case 91 Score = 77.6, SR = 6.94
ACTH Baseline (8:30AM) Day 2	Case 64 Score = 118. SR = 5.71
	Case 78 Score = 68.6 SR = 4.39
	Case 69 Score = 66.8 SR = 5.32
Cortisol Baseline Day 1	Case 15 Score = 29.5 SR = 4.23
Cortisol Baseline Day 2	Case 35 Score = 34.9 SR = 4.39

The apparent inconsistency between the SRs and Scores (i.e. the relative magnitudes of Scores and SRs are inverted) exists because we evaluated outliers at each stage of the analysis. That is, we initially assessed outliers with respect to the total distribution. After eliminating outliers based on studentized residuals for the total distribution we redid the outlier analysis on the new distribution. In this second stage, a score that did not appear as an outlier originally (because of the presence of more extreme scores in that distribution) might appear as an outlier in the new distribution (because that score is now the most extreme). However, the values of the studentized residuals are not comparable across the stages of the outlier analysis because they are based on different distributions.

Previous studies have suggested that baseline (8:30 AM) Cortisol levels may be high in individuals with FMS/CFS, whereas 24 hour urinary Cortisol may be normal or low (this discrepancy is felt to be due to disturbed circadian rhythmicity in CMI)[13]. For the baseline (8:30 AM) Cortisol on day one, we found a significant difference among the groups ($F_{2, 84} = 7.30$; p= .001) (see Figure 1 below)). The FMS/CFS subjects displayed higher mean Cortisol levels than Controls at all other points as well, that fell short of statistical significance at 3:30 PM (9.8 +/- 3.2 vs. 7.8 +/- 2.7; p=.07) and 8:30 AM day 2 (17.3+/- 6.0 vs. 13.1 +/- 4.1; p=.09) . The GWI subjects had levels between the FMS/CFS and Controls, that approached being significantly different from the Controls at 3:30 (9.7 +/- 3.9; p=.11) and 8:30 AM day 2 (16.9 +/- 5.6; p=.07). A follow-up test using a Bonferroni correction indicated that the mean of the FMS/CFS group (16.5) differed significantly from the mean of the Control group (12.3, p=.001) and from the mean of the gulf war group (13.2, p=.028).

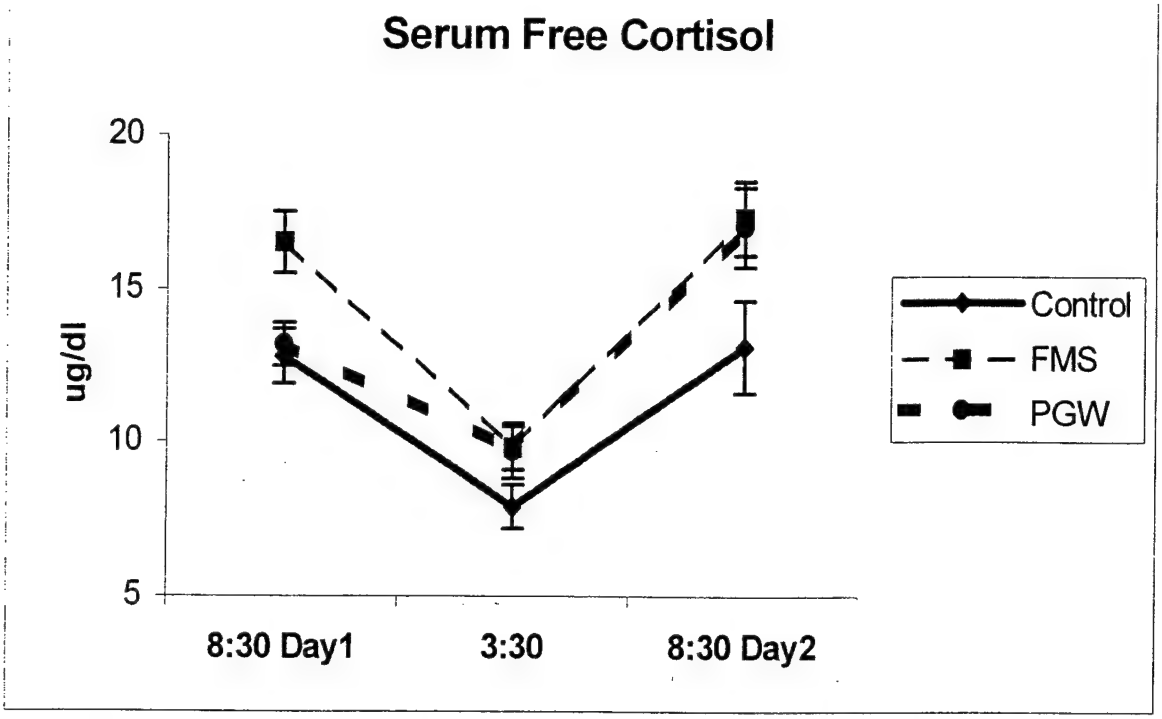


Figure 1 – Levels of serum free Cortisol the mornings of Day 1 and 2 and the afternoon of Day 1.

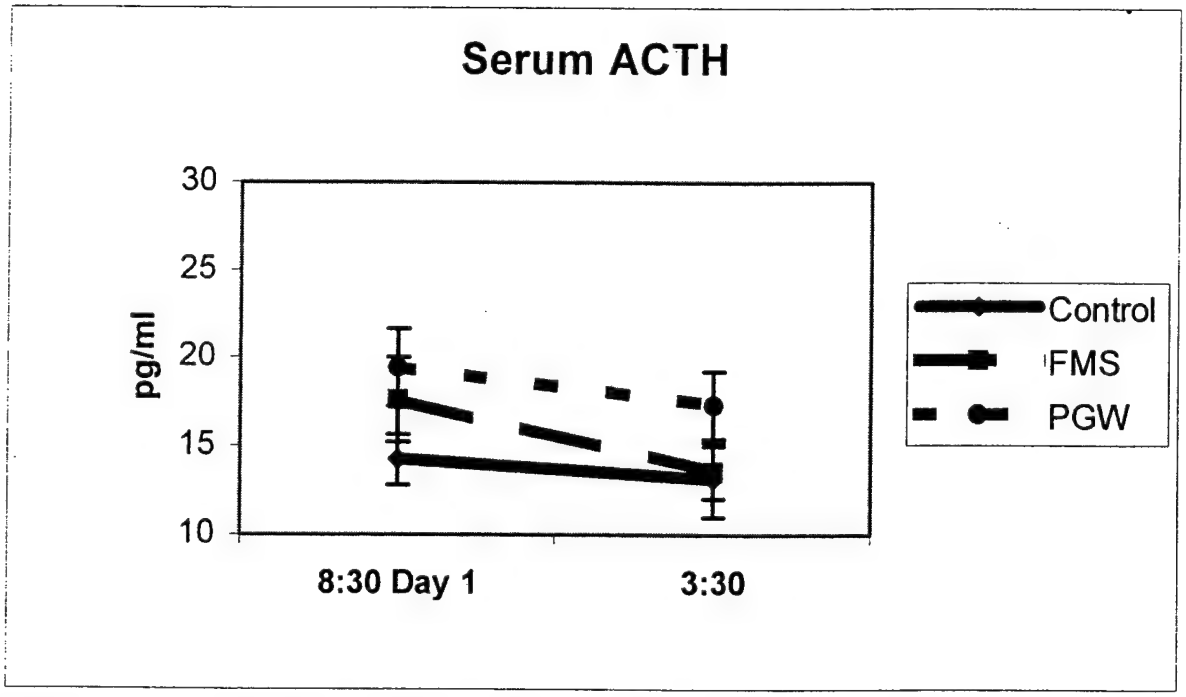


Figure 2 – Levels of serum ACTH on the morning and afternoon of Day 1

Although there were no significant differences between the GWI group and Controls for any of the Cortisol measures, Figure 1 demonstrates that the Cortisol levels of the GWI subjects for all points were between those of the FMS/CFS group, and Controls, and for the second two time points were virtually identical to the FMS/CFS group.

For the ACTH level at 8:30 AM and 3:30 PM, the group differences approached significance, ($F_{2, 88} = 2.54$, $p = .084$) (see Figure 2). A follow-up test using a Bonferroni correction suggested that the mean of the Control group (14.2) might differ from the mean of the GWI group (19.4; $p = .084$). Again, the values for the GWI group were much more similar to that of the FMS/CFS subjects, than Controls.

There were no statistically significant differences between groups found on the urinary Cortisol values. However, the 24 hour urinary Cortisol values collected on the day of exposure to stressors, Day 1 (Day 0 was collected on the day of admission to the CRC) again demonstrated that the GWI subjects fell between those of the FMS/CFS group and Controls (Figure 3).

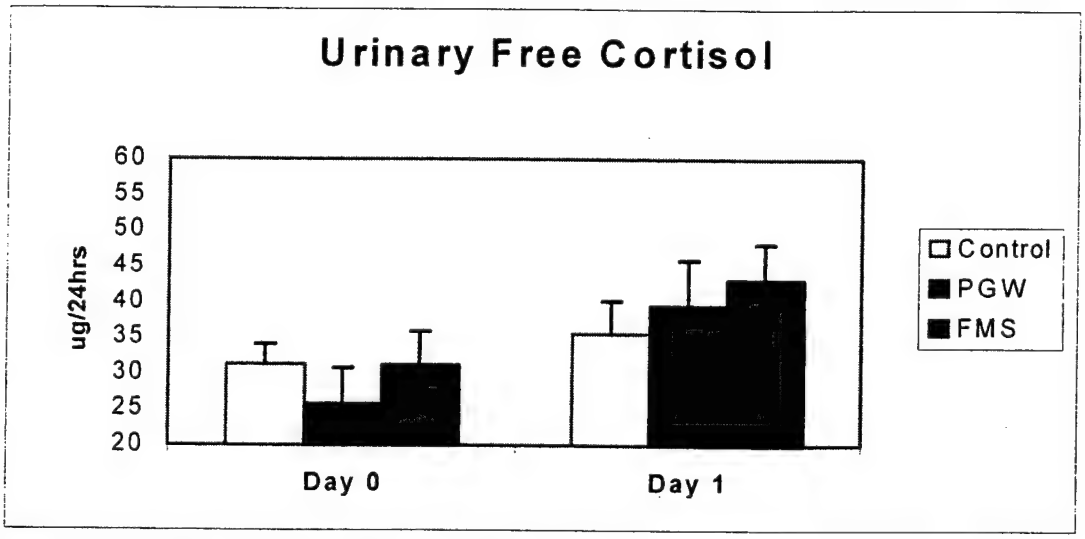


Figure 3 – Urinary free Cortisol on Days 0 and 1

The Issue of Outliers: In the previous analyses of Cortisol and ACTH we eliminated the outliers from the analyses because their inclusion substantially increased the size of the error term, thereby increasing the probability of a Type II error, especially in these analyses that are based on relatively small sample sizes. We only eliminated outlying subjects from the specific analysis on which they were outliers. Having said this, whether to keep or eliminate outliers is complex. There are at least four possible analytic strategies.

1) Keep all cases and do the analyses ignoring outliers.
2) Eliminate the outliers before doing the analyses (the strategy used here)
3) Transform the dependent variables to reduce the adverse impact of outliers on the statistical analyses, but keep the outliers in the data set.
4) Use a non-parametric version of the analysis of variance to eliminate the adverse effect of the outliers on the parametric assumptions of the analysis.

Strategies 1, 3 and 4 all assume that the outlying values, although extreme, are valid. That is,

they are not the result of measurement problems or methodological or data entry errors. Of these three strategies our preference is to transform the data to reduce the effect of the outlying values on the analysis, but keep those subjects in the data set. However we also did a non-parametric (Kruskal-Wallis) analysis of the group differences on the Cortisol and ACTH analyses.

Ignoring the outliers simply allows these few subjects to have too large an influence on the analysis, either by spuriously increasing the effect size or (as was the case here) inflating the within-group variability.

Thus we have redone the analyses comparing the groups with these two alternative treatments of the outliers.

We used a square root transformation of the Cortisol and ACTH variables to reduce the effect of the extreme scores on the analyses. Analyses of the square root transformed values for ACTH revealed no differences among the groups at any of the time points. This result was echoed by the non-parametric Kruskal-Wallis (KW) comparisons. Values of the KW test statistics and associated p values (based on a chi-square distribution with 2 degrees of freedom) were 3.73, $p = .16$, 2.49, $p = .29$, and 3.56, $p = .17$ for Baseline day 1, Baseline day 2, and Midday assessments, respectively.

For Cortisol, the analyses based on the square root transformed values and the results based on the non-parametric KW test were again similar to those obtained with the analyses where the outliers were eliminated. For Baseline Day 1 the analysis of variance of the square root transformed values was significant ($F_{2, 88} = 5.37$, $p = .006$) and the KW test statistic was 9.05, $p = .011$). For the baseline day 2 and midday assessments the groups were not found to differ on either the transformed values or on the KW test.

Thus, it does not appear that our removal of the outlying values in the original analysis seriously distorted our results. For reporting purposes we have decided to use the technique of eliminating outliers as described above.

Co-variates. As noted, there were several differences among groups in variables that might influence physiologic results. For sex, the GWI group had more men than women (18/25) whereas the FMS/CFS and Control groups consisted primarily of women (35/38, 43/49, respectively; chi-square with 2df = 40.57, $p = .000$). On the BDI (depression) there was also a significant difference among the groups ($p<.001$). Follow-up Bonferroni corrected tests indicated that the FMS/CFS (Mean = 13.7) and GWI (Mean = 13.4) groups displayed significantly more depressive symptomatology than the Control group (Mean = 4.1), but did not differ from each other (corrected probabilities both < .001). For the STAI (state anxiety) there again was a significant difference among the groups ($p=.000$). Again, follow-up Bonferroni corrected tests indicated that the FMS/CFS (Mean = 40.6) and GWI (Mean = 40.8) groups had significantly more anxiety symptomatology than the Control group (Mean = 28.5), but did not differ from each other (corrected probabilities both < .001).

Because of the strong associations between the groups and these variables we considered the differences among the groups on the baseline Cortisol and ACTH values, after statistically controlling for age, sex, depression (BDI), and anxiety (STAI). These are the two outcome variables for which we found some evidence of group differences in the uncontrolled analyses reported above.

We used a Hierarchical Linear regression for these analyses in which a control variable was entered as a predictor variable in step 1 of the analyses and then a dummy coded set of two variables to represent group membership was entered at the step 2. The unique contribution of group membership to predicting the dependent variable after controlling for the control variable is indexed by the squared semi-partial correlation of the set of variables representing group membership (i.e. the increment in the squared multiple R at step 2 over step 1).

For the baseline ACTH we no longer found evidence for group differences after controlling for age, sex, depression, or anxiety. However, for the baseline Cortisol variable the group differences generally remained significant after controlling for the variables. The results of these analyses are summarized in Table 8 below.

Table 8 – Analysis of covariance for 8:30 am serum free Cortisol

Control Variable*	squared semipartial	F	df	p
Age	.131	4.24	3, 84	.008
Sex	.113	3.58	3, 84	.017
Depression	.119	3.67	3, 78	.016
State Anxiety	.124	3.87	3. 80	.012
(all four)	.126	2.05	6, 75	.069

**The Depression variable is the index from the Beck Depression Inventory. The State Anxiety variable is the state anxiety index from the State/Trait Anxiety Inventory*

The squared semipartials can be interpreted as variance "accounted for". Thus, for example, the differences among the three groups account for 13.1% of the variance in baseline Cortisol at the first baseline assessment after controlling for age differences among the groups. The above squared semi-partials would generally be considered "large" (even "very large") effects.

Co-variates – II. We also considered that within both the FMS/CFS and GWI groups, physiologic variables might co-vary with the severity of the illness, based on self-reported pain, fatigue, physical functioning, as well as an overall severity index (combining all three of the above components). We then tested hypotheses about group differences on the severity indices, the role of the severity indices as covariates in the analyses of group differences on baseline Cortisol and ACTH, and the moderating effect of severity on the group differences in baseline Cortisol and ACTH. The composite variables were constructed using items from the McGill Pain questionnaire, the Multidimensional Fatigue Inventory, and the SF-36. the item-by-item analyses discussed earlier were performed after the analyses presented below.

A. Forming and Evaluating the Composites

The analyses below indicate the internal consistency reliability (coefficient alpha) and the range of the corrected item-total correlations for the four severity composites. (Note that the SF variables were reversed scored before they were included in the composites. Also to correct for differences in scaling the variables were all standard scored before creating the composites)

Pain Composite (McGill [MG] PAIN, MG Present Pain Intensity, MG VAS, MG Sensory, MG Affective, SF36 pain)

Coefficient Alpha - all items 0.954

Range of corrected item-total correlations: .841 to .898

Fatigue Composite (Multidimensional Fatigue Inventory [MFI]General Fatigue, MFI PF, MF IRA, MFI RM, MFI MF, SF36 fatigue scale)

Coefficient Alpha - all items 0.944

Range of corrected item-total correlations: .782 to .880

Functioning Composite (SF36 -01, SF36-03, SF36-08 functioning subscales)

Coefficient Alpha - all items 0.906

Range of corrected item-total correlations: .804 to .822

Overall Severity Composite (all of the above 15 items)

Coefficient Alpha - all items 0.965

Range of corrected item-to-total correlations: .656 to .893

All of the values of coefficient alpha are substantial indicating that the composites provide highly reliable and representative indices of pain, fatigue, physical functioning and overall severity. The range of the corrected item to total correlations suggest that the individual variables that make up the composites are all consistent indicators of the construct assessed by the composite. Post hoc group comparisons of all pain and fatigue variables were performed and are presented above.

B. Group differences of the severity composites

We compared the means of the FMS/CFS, GWI, and Control groups on each of the four composites in four one-way analyses of variances. The results are summarized in Table 9 below. The negative values for the Control group means reflect how the composite severity index was formed. Because the variables that were used to index overall severity are on different scales we first standard ("z") scored the variables before averaging them. As indicated (and expected), there are substantial differences between the Control group and the FMS/CFS and Gulf War Groups. Bonferroni corrected follow-up comparisons among the group means indicate significant differences between the Control group mean and the means of the FMS/CFS and Gulf War Groups on all the composites. However we found no significant differences between the Gulf War and FMS/CFS. Note that these unweighted composite scores can be interpreted as approximate standard (z) scores. Thus, for example the self-reported fatigue of the FMS/CFS and Gulf War groups is roughly 3/4 of a standard deviation above the mean self reported fatigue of all subjects whereas the Control group self-reported fatigue is roughly 3/4 of a standard deviation below that mean.

Table 9 - Comparison among groups of standardized severity composites.

Group Means

Composite Variable	FMSCFS	GWI	Control	F	P
Fatigue	.74	.73	-.70	73.8	<.0001
Pain	.56	.74	- .79	74.0	<.0001
Physical Functioning	.62	.50	-.97	139.8	<.0001
Overall Severity	.66	.67	-.82	136.2	<.0001

The ranges for the total sample are:

Fatigue	-1.52 to 1.96
Pain	-1.30 to 1.81
Physical Functioning	-1.25 to 1.48
Overall Severity	-1.32 to 1.48

The composite scores are highly consistent with the post-hoc analyses for individual measures.

C. The effect of using the severity indices as co-variates in the analyses to compare the groups on baseline Cortisol and ACTH.

We restricted our attention to the baseline Cortisol and ACTH outcome measures, as these are the two outcomes that consistently showed some differences. These analyses were again done after removing outliers as described earlier. The results are summarized in Tables 10 and 11 below. Again the group differences for ACTH generally disappear when any of the severity covariates are included, whereas for Cortisol the group differences remain.

Each squared semipartial is the proportion of variance in outcome accounted for uniquely

by group membership after controlling for differences in severity.

Table 10 – Squared semipartials for covariates of ACTH

Control Variable	Squared semipartial	DF	F	P
Fatigue	.064 - .006 = .058	2, 81	2.51	.09
Pain	.060 - .006 = .054	2, 83	2.38	.10
Physical function	.050 - .022 = .028	2, 83	1.22	.30
Severity	.050 - .018 = .032	2, 84	1.41	.25
Total*	.144 - .132 = .012	2, 71	.50	.61

*Total refers to pain, fatigue, and physical functioning plus all the earlier covariates–Depression, Gender, Anxiety, and Age

Table 11 – Squared semipartials for covariates of Cortisol

Control Variable	Squared semipartial	DF	F	P
Fatigue	.216 - .107 = .109	2, 76	5.28	.0071
Pain	.206 - .024 = .182	2, 78	8.94	.0003
Physical function	.186 - .104 = .082	2, 78	3.93	.02
Severity	.184 - .075 = .109	2, 79	5.28	.007
Total	.278 - .202 = .076	2, 66	3.47	.0369

D. The Effect of Severity as a Moderator of the Group Differences on ACTH and Cortisol

The final analyses we report here address the question: does self-reported severity relate to the physiological outcome measures differently for the different groups? To answer this question we ran hierarchical regression analyses in which we first entered the dummy coded group variables and the severity index as predictors of either Cortisol or ACTH. We then entered the products of the severity index and the two dummy coded variables. The increment in the R^2 from adding the product over the R^2 from the regression equation containing the dummy variables and the severity index indicates the effect size (squared semi-partial) of the interaction of group and severity for predicting outcome. The relevant models are summarized in Table 12 below. Only the results for Cortisol are summarized as no interaction was found for ACTH.

Table 12 - Summary of the Hierarchical Regression analysis to test the interaction of group and severity for predicting Cortisol

		Increment in R^2			
Step	**Variables entered**	**sr^2**	**F for increment**	**df**	**p**
1	Group	.15	7.30	2, 84	<.001
2	Severity	.03	< 1.0	1, 79	.92
3	Group x Severity	.06	3.75	2, 77	.028

The group x severity interaction for predicting Cortisol accounted for 6% of the variability in Cortisol, which represents a moderate effect size. This effect was also statistically significant.

Significance: These data replicate some of the previous studies demonstrating differences in hypothalamic pituitary adrenal function between individuals with FMS/CFS and Controls. The main positive finding in this study was that the FMS/CFS group had a higher morning Cortisol than the Controls; this finding remained significant when all co-variates were considered. Other studies of CMI have detected similar differences between patients and Controls (13), especially when individuals are studied under conditions of "stress". One unique aspect of our studies, when compared to some of the previous studies done examining the HPA axis in these illnesses, is that our subjects had these tests performed while engaging in identical activities, in the regulated environment of a clinical research center.

Although we did not find that the GWI subjects had precisely the same changes as the FMS/CFS subjects, *on nearly all neuroendocrine measures the GWI cohort had values that were either intermediate between normals and individuals with FMS/CFS, or even more abnormal than those of the FMS/CFS groups* (this phenomena is best depicted on Figures 1, 2, and 3). This is seen repeatedly in the data collected in this study, and is likely due to the fact that the "criteria" for GWI identify individuals with GWI who suffer from a "sub-syndromal" form of FMS and/or CFS, and do not meet the strict criteria for either of these illnesses, but approach those criteria. The fact that for most time points these findings approached statistical significance, but did not reach statistical significance, suggests that future studies of these phenomena should include larger numbers of subjects, and examine more time points to estimate the circadian rhythmicity of HPA axis function.

Autonomic function.

Tilt table testing. Using the protocol described, we examined tilt-table variables (systolic blood pressure=SBP, diastolic blood pressure=DBP, and heart rate=HR) to see (1) if there were any differences in HR, SBP, or DBP among groups and across time, and (2) to look for group differences in the time of syncope/presyncope onset.

The variables of interest were:

TILTDOWN: 1 = yes–syncope/presyncope 2 = no

SBP by time: SYS0, SYS5, SYS10, SYS15, SYS20, SYS25, SYS30
Time 0 is baseline, and all other times are changes in SBP compared to baseline.

DBP by time: DIA0, DIA5, DIA10, DIA15, DIA20, DIA25, DIA30

HR by time: HR0, HR5, HR10, HR15, HR20, HR25, HR30

Measures taken 10 min after the test: POSTSYS, POSTDIA, POSTHR

1) The first question we addressed was whether or not there were any group differences in developing syncope/presyncope. Specifically, we examined how fast subjects developed syncope and if there were differences in groups. We did this by calculating the number (and %) of people in each group that were left at each time period.

Table 13 below contains values for the number of subjects in each group who were remaining at each time, and the percent of the original group that the number reflects.

Table 13

Group	0	5	10	15	20	25	30
Control (n = 49)	49	49	46	39	37	32	27
% remaining	100	100	94	80	75	65	55
FMS/CFS (n = 36)	36	34	31	28	25	24	23
% remaining	100	94	86	78	69	67	64
GWI (n = 24)	24	23	22	20	19	19	19
%remaining	100	96	92	83	79	79	79

From just "eye-balling" this table, it appears that a larger % of the GWI subjects did not develop syncope on tilt, but the differences are not large.

2) To examine group differences in SBP, DBP, and HR over time, we conducted three mixed-model ANOVAs. Each was a 3 (group) x 6 (time) ANOVA with repeated measures on the second factor (time). The dependent variables were SBP, DBP, and HR. For the HR and SBP analyses, there were no significant effects for group, time, or the group x time interaction. However, for DBP, there was a main effect for group ($F(2,65) = 4.29$, $p = .018$) and a main effect for time ($F(5,325) = 2.70$, $p = .021$). Figure 4 shows the changes in DBP with time for each group The main effect for group appears to be caused by fact that the GWI group has the

largest increases in DBP at five of the six points in time. As for the time effect, there seems to be a general tendency in the Control and FMS/CFS groups for DBP to remain steady or fall over

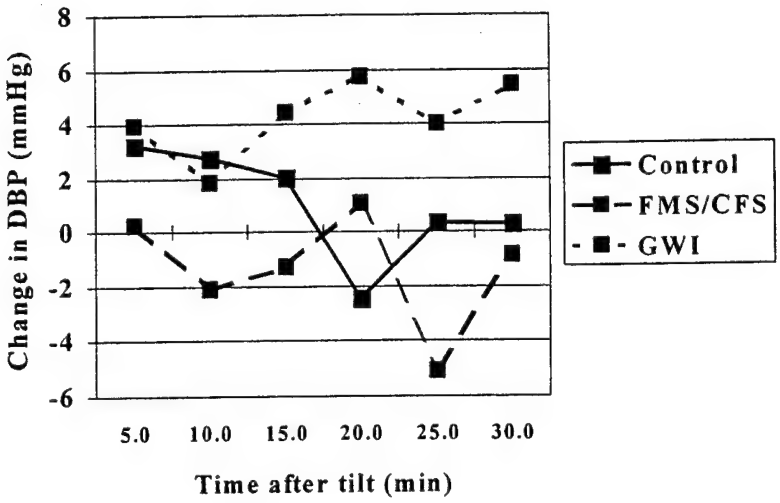
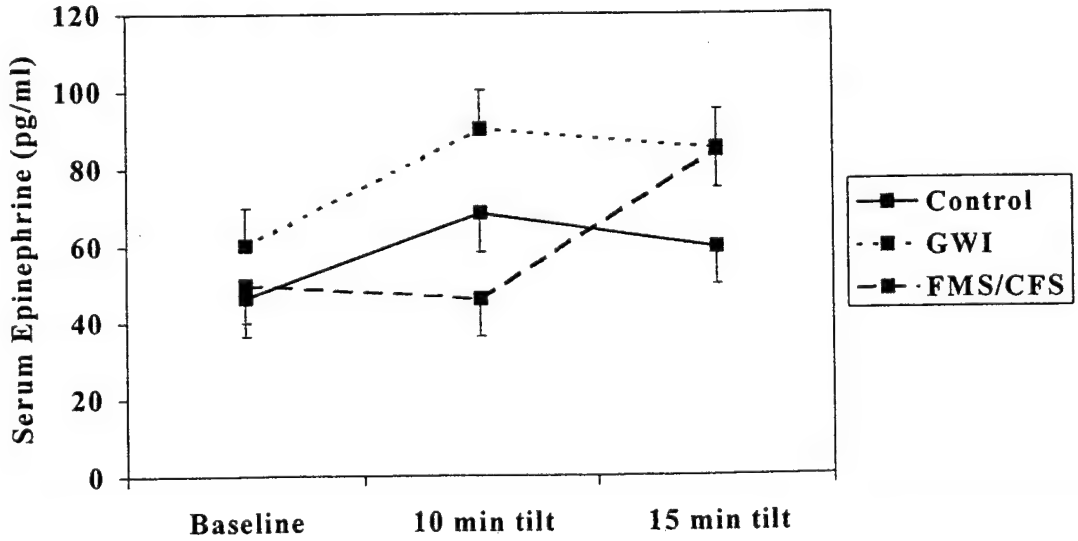


Figure 4 – Change in DBP over time during upright tilt

time. The values for the GWI group actually go up over time, but not enough for the interaction term to be significant.

Catecholamine responsiveness to tilt table testing. In addition to assessing whether individuals had a "positive" or "negative" tilt table test, we also examined the release of catecholamines in response to orthostasis. Figures 5 and 6 show the values for epinephrine and norepinephrine at baseline (20 min prior to tilt) and at 10 and 15 min into tilt.



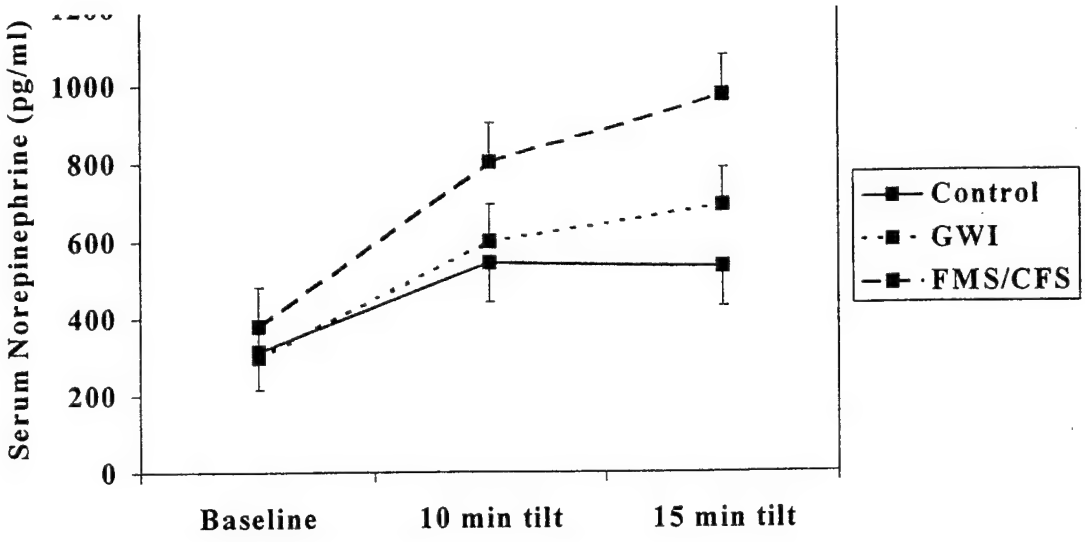


Figure 5 – Epinephrine levels as a function of time during tilt

Figure 6 – Norepinephrine levels as a function of time during tilt

A significant problem with the epinephrine values in this study is that a large number of samples had undetectable levels (the reference laboratory that we used for this study had a lower limit of detection of 40 pg/ml), and for these samples we arbitrarily entered a value of 40. Because of this, it is difficult to make any meaningful inferences from these values. However, using a repeated measures anova we found (Figure 5) that time had a significant effect ($p=0.01$, indicating a response to tilt) and there was a marginal group effect ($p=0.152$). Control values were lower than GWI ($p=0.064$), however there was no difference between Controls and FMS/CFS. Even though epinephrine levels were higher in the GWI group compared to the Control group, the pattern of epinephrine response appeared similar.

There was no measurement problem for the norepinephrine (Figure 6). As with epinephrine there was a significant time effects ($p=0.000$, indicating a response to tilt) and a marginally significant group effect ($p=0.110$). In contrast to epinephrine, norepinephrine was significantly higher in FMS/CFS compared with Controls ($p=0.038$). There were no differences between Control and GWI groups or between GWI and FMS/CFS groups. These data suggest that the norepinephrine response may be accentuated in FMS/CFS subjects, but possibly delayed. This contrasts with an epinephrine response that was no different from Control. Whether the response differences between FMS/CFS and GWI subjects mean anything is difficult to determine because the subject numbers were small (n=6 FMS/CFS; n=10 GWI; and n=12 Control) and the effect of the epinephrine measurement problems indeterminate. However, if we take the data at face value, they suggest that there is little difference in the pattern of epinephrine response among all three groups, whereas FMS/CFS subjects may have a norepinephrine hyperresponse to tilt,
with GWI falling between Control and FMS/CFS. These data may explain the higher DBP in the GWI subjects during tilt and may have played a role in actually preventing syncope. In any event, there is little evidence here to suggest that either GWI or FMS/CFS subjects are hyporesponseive to orthostatic stress.

Since the time that this grant was originally written, several groups including ours has questioned whether tilt table testing is more likely to be abnormal in CMI patients than Controls (in contrast to the original reports from Johns Hopkins) [14-16]. Although we observed some differences among groups from both catecholamine and hemodynamic response as described above, these responses are inconsistent with earlier reports. The problem may be related to inherent problems with the tilt table test. First, prior to tilt some subjects may have alterations in autonomic function in anticipation of being tilted. Second, in our hands it was difficult to ensure that all subjects maintained a relaxed posture during tilt. For this reason, our ongoing studies of CMI subjects have eliminated tilt table testing, and have substituted exercise testing, which has been consistently found to be an adequate "stressor", and lead to differences in autonomic responsiveness between groups (17).

Heart rate variability. Heart rate variability (HRV) monitoring has been demonstrated to be an accurate means of assessing both the parasympathetic component heart rate variability, and can approximate the sympathetic component. Heart rate variability can be expressed in either the time domain or in the frequency domain. In the frequency domain, the low frequency (LF) component of HRV (0.04-0.15 Hz) is roughly equal to the cardiac sympathetic input to heart rate variability (although it may have a parasympathetic component), whereas the high frequency (HF) component of HRV (0.14-0.4 Hz) is an accurate measure of cardiac parasympathetic component. The ratio LF /HF is thought by some to be a measure of sympathovagal balance.

Measures of HRV in the time domain include:

AVNN	Average time interval (N-N) in msec between normal QRS complexes
SDNN	Standard deviation of all N-N intervals in msec
SDANN	Standard deviation of the 5-minute averages of N-N intervals in msec measured over a recording period
SDIDX	Standard deviation of differences in msec between adjacent N-N intervals
PNN50	Percent of all N-N intervals for which the number of pairs of adjacent N-N intervals differs by more than 50 ms over a recording period

Measures of HRV in the frequency domain include:

TP	Total power in $msec^2$ of the entire frequency distribution
LF	Power in $msec^2$ in the LF range
HF	Power in $msec^2$ in the HF range
LF(nu)	Low frequency power divided by (TP-VLF)
HF(nu)	High frequency power divided by (TP-VLF)
LF/HF	Ratio of LF/HF

Heart rate variability (HRV) was analyzed over continuous five minute intervals in the time and frequency domains for 24-hours (T), daytime (D: 08:00-20:00) and nighttime (N: 00:00-06:00). Un-paired t-tests were calculated separately for each group pair: GWI vs. Control, FMS/CFS vs. Control, and GWI vs. FMS/CFS.

Analysis

Variables in the Analyses

In this analysis there are 33 HRV variables of interest. For each of the following variables, there are 3 measurements (24-hour, Daytime, Nighttime):

AVNN
PNN50
RMSSD
SDNNT
SDIDX
SDANN
PNN625
TP5
LFHF5
LF5(nu)
HF5(nu)

Distributional Considerations.

HRV data tends to be skewed and to contain extreme, but valid cases. Including such cases in a statistical analysis can have an adverse impact on the analysis because such cases may magnify effects that may not replicate (Type I error) or inflate an error term and therefore obscure effects (Type II error). Unfortunately, excluding such cases means that legitimate observations are removed from the data set for purely statistical reasons. Out of a concern for these issues we have adopted two analytic strategies. One is to log transform the HRV variables which may reduce the effect of extreme values on the statistical conclusions without removing those observations from the analyses. Log transformation is a standard practice in the evaluation of HRV data. For the LF and HF components of power spectral density, these values were normalized for each subject to (TP-VLF) and expressed as a percent. This effectively removes influence of the VLF component on LF and HF. Normalized LF and HF are indicated as LF(nu) and HF(nu) respectively.

The other strategy is to use a distribution-free test to compare the groups on the HRV variables. In this case the Kruskal-Wallis test, in which the scores of the participants on each dependent variable are transformed to ranks across all groups, is used. The rank-ordered data is then divided into the groups (in this case GWI, FMS/CFS, and control) and the centers of the ranked scores are compared between the groups. The null hypothesis that is tested is that the center of the ranks is the same for the groups. This comparison is a distribution free analog to the one-way analysis of variance. No assumptions are made that the distribution is normal, which is consistent with the non-normal (i.e. skewed) nature of HRV data. One limitation of the Kruskal-Wallis test is that covariates and more complex designs can not be evaluated. However, significant group differences with the Kruskal-Wallis test that are consistent with group differences on more conventional tests provide some indication that the conventional results are not an artifact of their distribution.

Group Analysis

Each of the 33 variables was log transformed to reduce the impact of extreme values on the analyses without eliminating the extreme but valid cases from the analyses. Table 14 below contains means and standard deviations for each variable. P-values are also included for each 1 x 3 ANOVA (p_1) and each 1 x 3 Kruskal-Wallis Test (p_2). For each significant effect, more detailed results are provided later in this report. Note: For 24-hour variables, subjects with HRVTOK = 0 were not included. For daytime, HRVDOK = 0 were excluded. For nighttime, HRVNOK = 0 were excluded. In addition, all medicated FMS/CFS subjects were removed. It should also be noted that these analyses were conducted on ALL subjects with HRV data. As will be described later in this report, the covariate analyses had to be conducted with a smaller subset of subjects because some of the subjects had HRV data, but no covariate data (BDI, composite variables, etc.)

Table 14 Group comparisons for HRV data using Anova and Kruskal-Wallis Test

Variable	All (N = 92)	Control (N = 40)	FMS/CFS (N = 35)	GWI (N = 17)	p_1	p_2
AVNNT	766.5 ± 97.1	796.9 ± 106.0	735.56 ± 74.0	758.9 ± 100.5	.02	.03
AVNND	728.0 ± 95.8	755.1 ± 104.8	699.9 ± 67.9	724.0 ± 111.1	.04	.05
AVNNN	861.7 ± 117.2	896.6 ± 129.4	823.5 ± 102.6	860.6 ± 92.8	.01	.05
PNN50T	9.8 ± 11.5	12.3 ± 12.9	6.0 ± 8.3	11.5 ± 12.1	.03	.03
PNN50D	8.2 ± 10.3	10.9 ± 11.9	4.8 ± 7.0	9.3 ± 10.8	.02	.03
PNN50N	12.8 ± 15.3	15.3 ± 17.5	7.9 ± 10.6	17.5 ± 16.3	.03	.04
RMSSDT	33.2 ± 23.6	38.0 ± 26.9	26.9 ± 20.1	35.2 ± 20.1	.03	.02
RMSSDD	30.7 ± 21.5	35.4 ± 23.6	25.0 ± 19.2	32.0 ± 18.5	.03	.03
RMSSDN	37.4 ± 30.0	42.1 ± 37.2	29.7 ± 21.1	42.8 ± 24.4	.03	.04
SDNNT	120.9 ± 37.3	133.7 ± 44.4	107.8 ± 27.0	118.0 ± 27.6	.02	.03
SDNND	103.1 ± 38.4	116.8 ± 46.3	91.1 ± 24.4	96.3 ± 33.0	.02	.07
SDNNN	89.03	95.84	78.14	96.32	.03	.04
SDIXDT	56.7 ± 24.3	62.7 ± 29.6	48.7 ± 16.6	59.4 ± 19.7	.04	.06
SDIXDD	56.28 ± 23.4	62.6 ± 28.4	48.4 ± 15.8	58.3 ± 19.9	.03	.07
SDIXDN	57.7 ± 28.1	62.5 ± 35.2	49.8 ± 19.2	63.2 ± 21.8	.05	.09

SDANNT	104.0 ± 32.1	114.8 ± 37.8	93.9 ± 25.4	99.7 ± 21.6	.03	.06
SDANND	84.9 ± 34.5	97.7 ± 41.6	75.1 ± 22.6	75.3 ± 27.5	.01	.03
SDANNN	60.2 ± 23.4	64.0 ± 24.6	54.3 ± 19.3	63.9 ± 27.0	.14	.16
PNN625T	9.3 ± 9.8	11.1 ± 10.8	6.4 ± 8.0	11.2 ± 9.8	.04	.04
PNN625D	8.8 ± 9.2	10.8 ± 10.5	6.0 ± 7.2	9.9 ± 8.4	.04	.04
PNN625N	10.3 ± 12.5	11.7 ± 13.8	6.6 ± 9.4	14.9 13.4	.04	.06
TPT5	4208.2 ± 4111.2	5252.6 ± 5313.3	2977.1 ± 2449.9	4285.1 ± 2917.8	.05	.07
TPD5	4066.7 ± 3825.4	5097.2 ± 4825.1	2894.0 ± 2043.8	4133.8 ± 2976.6	.05	.06
TPN5	4424.1 ± 5143.3	5442.0 ± 7027.2	3166.8 ± 2612.2	4705.8 ± 3196.4	.06	.08
LFHFT5	4.5 ± 2.3	3.8 ± 1.8	5.3 ± 2.9	4.2 ± 1.6	.13	.07
LFHFD5	5.1 ± 2.7	4.4 ± 2.2	6.0 ± 3.3	4.7 ± 1.7	.14	.09
LFHFN5	3.5 ± 2.3	3.0 ± 1.8	4.4 ± 2.8	3.0 ± 1.5	.09	.04
LFT5(nu)	55.0 ± 8.7	52.4 ± 7.4	56.9 ± 9.6	57.3 ± 8.3	.06	.03
LFD5(nu)	56.1 ± 8.5	53.2 ± 6.7	57.5 ± 9.7	60.1 ± 7.8	.01	.007
LFN5(nu)	54.5 ± 11.4	52.2 ± 11.5	57.5 ± 11.3	53.8 ± 10.2	.24	.08
HFT5(nu)	21.4 ± 9.1	22.9 ± 8.4	19.6 ± 10.6	22.4 ± 6.7	.07	.06
HFD5(nu)	17.6 ± 8.2	18.9 ± 7.5	16.4 ± 9.6	17.3 ± 6.0	.15	.10
HFN5(nu)	27.7 ± 12.3	30.1 ± 12.6	24.0 ± 12.2	29.8 ± 10.4	.04	.03

Post hoc tests for group comparions

Of the 33 HRV variables, there were statistically significant differences ($p < .05$) on 24 using the ANOVA results. Table 15 below contains data from the statistically significant Bonferroni corrected post hoc tests. D = mean difference. Remember, that thee analyses were conducted with log-transformed variables, so mean differences are in log transformed units.

Table 15 – Pairwise comparisons of FMS/CFS, GWI, and Control Groups

24-hour		
AVNNT	Control > FMS/CFS	(D = .077, p = .018)
PNN50T	Control > FMS/CFS	(D = .760, p = .054)
RMSSDT	Control > FMS/CFS	(D = .306, p = .035)
SDNNT	Control > FMS/CFS	(D = .194, p = .015)
SDIDXT	Control > FMS/CFS	(D = .208, p = .058)
SDANNT	Control > FMS/CFS	(D = .187, p = .022)
PNN625T	(no follow up tests with p < .08)	
TPT5	Control > FMS/CFS	(D = .394, p = .072)
LFHFT5	F was not statistically significant (see Table of means above)	
LFT5(nu)	F was not statistically significant (see Table of means above)	
HFT5(nu)	F was not statistically significant (see Table of means above)	
Daytime		
AVNND	Control > FMS/CFS	(D = .071, p = .034)
PNN50D	Control > FMS/CFS	(D = .821, p = .032)
RMSSDD	Control > FMS/CFS	(D = .309, p = .026)
SDNND	Control > FMS/CFS	(D = .214, p = .018)
SDIDXD	Control > FMS/CFS	(D = .215, p = .035)
SDANND	Control > FMS/CFS	(D = .231, p = .017)
PNN625D	Control > FMS/CFS	(D = .583, p = .078)
TPD5	Control > FMS/CFS	(D = .402, p = .057)
LFHFD5	F was not statistically significant (see Table of means above)	
LFD5(nu)	GWI > Control	(D = .122, p = .019)
HFD5(nu)	F was not statistically significant (see Table of means above)	
Nighttime		
AVNNN	Control > FMS/CFS	(D = .091, p = .008)
PNN50N	GWI > FMS/CFS	(D = 1.202, p = .037)
RMSSDN	Control > FMS/CFS	(D = .297, p = .072) & GWI > FMS/CFS (D = .391, p = .063)
SDNNN	Control > FMS/CFS	(D = .194, p = .058)
SDIDXN	(no follow up tests with p < .09)	
SDANNN	F was not statistically significant (see Table of means above)	
PNN625N	GWI > FMS/CFS (D = 1.08, p = .034)	
TPN5	F was not statistically significant (see Table of means above)	
LFHFN5	F was not statistically significant (see Table of means above)	
LFN5(nu)	F was not statistically significant (see Table of means above)	
HFN5(nu)	Control > FMS/CFS	(D = .240, p = .055)

With few exceptions, post hoc tests showed that the differences were between the FMS/CFS and Control groups. For each of the daytime time-domain variables the FMS/CFS had significantly lower values compared to controls with GWI subjects falling between the Control and FMS/CFS groups, clearly indicating that the FMS/CFS group had lower HRV compared to Control with a non-significant trend for the GWI group. For the daytime frequency domain variables TP mirrored the findings with the time-domain variables. However, only LF showed a group effect during the day with the GWI group having significantly higher LF compared to Control. In several instances, the means of the HRV variables for the GWI group were very close to those of the FMS/CFS group, but these differences were usually not statistically

significant, perhaps because of the small N for GWI (N = 17). For HRV variables during nighttime, many of the differences observed during the day disappeared with the GWI group looking more like the Control group. Of note, however, is that PNN50, and PNN625 in the GWI group became significantly higher than FMS/CFS group. In addition, group differences in the frequency domain variables disappeared with the exception of HF, which was significantly lower than either Control or GWI. We will later discuss day/night differences.

Comparisons Based on only those subjects for whom Covariate Information is Available

Because several subjects that were included in the above analyses did not have data for the covariate variables. A separate set of ANOVAs was calculated using only the subjects who had data on both the covariate and HRV variables. This dataset is, of course, smaller. Table 16 below contains the means and standard deviations. The results are similar to those provided in Table 14 above. The data are presented in non-log-transformed units. However, ANOVAs and KW statistics were computed using log transformed data. As before, these analyses were run with HRVTOK = 1 or HRVDOK = 1 or HRVNOK = 1 (depending on the time of measurement).

Table 16 Group comparisons for HRV data in subjects with co-variate data using Anova and Kruskal-Wallis Test

Variable	All (N = 71)	Control (N = 31)	FMS/CFS (N = 23)	GWI (N = 17)	p-value ANOVA	p-value KW
AVNNT	774.0 ± 97.6	804.4 ± 105.1	744.2 ± 74.2	759.0 ± 100.5	.06	.07
AVNND	736.8 ± 99.8	768.3 ± 107.3	703.7 ± 65.7	724.0 ± 111.1	.05	.05
AVNNN	867.0 ± 112.5	893.8 ± 123.5	835.5 ± 105.4	860.6 ± 92.8	.17	.22
PNN50T	10.9 ± 12.3	13.7 ± 13.6	6.7 ± 9.5	11.5 ± 12.1	.06	.06
PNN50D	9.6 ± 11.2	12.8 ± 12.5	5.6 ± 8.3	9.3 ± 10.8	.04	.04
PNN50N	14.2 ± 16.6	16.6 ± 18.7	8.7 ± 12.5	17.5 ± 16.3	.07	.08
RMSSDT	35.6 ± 25.9	40.9 ± 29.2	28.6 ± 24.2	35.2 ± 20.1	.07	.04
RMSSDD	33.6 ± 23.7	39.3 ± 25.4	26.9 ± 23.8	32.0 ± 18.5	.04	.03
RMSSDN	39.9 ± 33.5	44.7 ± 41.4	31.4 ± 26.1	42.8 ± 24.4	.10	.09
SDNNT	123.4 ± 38.4	134.4 ± 46.6	112.5 ± 29.3	118.0 ± 27.6	.19	.17
SDNND	108.0 ± 41.4	124.0 ± 49.0	95.2 ± 27.1	96.3 ± 33.0	.03	.06
SDNNN	92.1 ± 35.4	99.1 ± 42.2	79.5 ± 25.2	96.3 ± 30.5	.16	.17
SDIXDT	59.2 ± 25.8	65.5 ± 31.4	50.6 ± 18.9	59.4 ± 19.7	.15	.15

SDIXDD	55.6 ± 25.1	66.7 ± 29.8	50.9 ± 18.6	58.3 ± 19.9	.10	.10
SDIXDN	60.0 ± 30.2	64.7 ± 38.0	51.2 ± 21.6	63.2 ± 21.8	.20	.30
SDANNT	105.1 ± 32.2	113.1 ± 39.0	98.3 ± 26.9	99.7 ± 21.6	.31	.40
SDANND	88.8 ± 37.1	103.6 ± 44.1	78.7 ± 24.9	75.3 ± 27.5	.02	.02
SDANNN	62.1 ± 23.8	66.6 ± 26.0	54.6 ± 16.1	63.9 ± 27.0	.35	.20
PNN625T	10.3 ± 10.6	12.4 ± 11.6	6.7 ± 9.4	11.2 ± 9.8	.06	.07
PNN625D	10.0 ± 9.9	12.6 ± 11.1	6.5 ± 8.6	9.9 ± 9.8	.06	.05
PNN625N	11.5 ± 13.8	13.0 ± 15.2	7.1 ± 11.4	14.9 ± 13.4	.05	.07
TPT5	4597.6 ± 4450.5	5748.0 ± 5725.0	3277.9 ± 2894.3	4285.1 ± 2917.6	.17	.17
TPD5	4563.8 ± 4166.4	5757.6 ± 5160.1	3272.7 ± 2929.9	4133.8 ± 2976.6	.13	.07
TPN5	4825.0 ± 5709.0	5916.5 ± 7814.0	3442.0 ± 3053.4	4705.8 ± 3196.4	.23	.29
LFHFT5	4.5 ± 2.4	3.8 ± 1.8	5.8 ± 3.2	4.2 ± 1.6	.10	.06
LFHFD5	5.0 ± 2.7	4.2 ± 2.2	6.4 ± 3.4	4.7 ± 1.7	.09	.04
LFHFN5	3.6 ± 2.4	3.0 ± 1.9	4.7 ± 3.1	3.0 ± 1.5	.06	.04
LFT5(nu)	55.2 ± 9.0	52.5 ± 8.0	57.4 ± 10.2	57.3 ± 8.3	.12	.06
LFD5(nu)	56.5 ± 8.9	53.3 ± 7.2	58.2 ± 10.4	60.1 ± 7.8	.03	.02
LFN5(nu)	54.3 ± 11.8	52.1 ± 12.4	57.7 ± 11.7	53.8 ± 10.2	.26	.17
HFT5(nu)	21.3 ± 9.5	23.3 ± 8.8	18.7 ± 11.6	21.4 ± 6.7	.04	.02
HFD5(nu)	17.8 ± 8.6	19.6 ± 7.9	15.8 ± 10.9	17.3 ± 6.0	.04	.02
HFN5(nu)	27.6 ± 12.7	30.2 ± 13.4	22.6 ± 12.9	29.8 ± 10.4	.02	.02

Post hoc tests on Group Comparisons Of the 33 HRV variables, there were statistically significant ($p < .05$) for 10 using the ANOVA results. With few exceptions, post hoc tests showed that the differences were between the FMS/CFS and Control groups. Again, in several instances, the mean for the GWI group was very close to that of the FMS/CFS group, but the differences compared to Controls were not statistically significant mostly because of the small N for GWI (N = 17). Table 17 below contains data from the statistically significant Bonferroni

corrected post hoc tests. D = mean difference. Remember, that thee analyses were conducted with log-transformed variables, so mean differences are in log transformed units.

Table 17 – Results of post-hoc comparisons for variables with statistically significant Bonferroni corrected tests.

24-hour					
NHFT5	Control > FM	(D = .294, p = .033)			
Daytime					
AVNND	Control > FM	(D = .083, p = .054)			
PNN50D	Control > FM	(D = .997, p = .038)			
RMSSDD	Control > FM	(D = .380, p = .038)			
SDNND	Control > FM	(D = .232, p = .060)			
SDANND	Control > FM	(D = .243, p = .062)	&	Control > GWI	(D = .286, p = .040)
LFD5(nu)	GWI > Control	(D = .121, p = .040)			
HFD5(nu)	Control > FM	(D = .299, p = .037)			
Nighttime					
PNN625N	GWI > FM	(D = 1.20, p = .046)			
NHFN5	Control > FM	(D = .319, p = .032)	&	GWI > FM	(D = .340, p = .055)

Despite the reduction in N, the results of statistical analysis demonstrate the same relationships among variables as were found with the larger dataset.

Analyses controlling for the Covariates

Covariate analyses were carried out for each of these 10 variables. That is, for each variable, an analysis was undertaken to determine if the group effect remained after the effects of the covariate(s) had been taken into account. The significance tests in Table 18 below are for the effect of group one the specified covariate has been taken into account.

Table 18 – Results from covariate analyses for variables showing significant group effects.

NHFT5

Control Variable	Squared semipartial	DF	F	P
1. Age	.188 - .068 = .024	2, 67	4.95	.01*
2. BDI	.112 - .076 = .036	2, 64	1.30	.28
3. STAI	.253 - .209 = .044	2, 65	1.91	.16
4. Gender	.144 - .014 = .130	2, 67	5.09	.008*
5. Fatigue	.085 - .050 = .035	2, 65	1.24	.29
6. Pain	.144 - .099 = .045	2, 66	1.73	.18
7. Physical function	.095 - .053 = .042	2, 65	1.51	.23
8. Severity	.120 - .082 = .038	2, 66	1.43	.25
9. Total (1-7)	.475 - .451 = .024	2, 55	1.26	.29

AVNND

Control Variable	Squared semipartial	DF	F	P
1. Age	.077 - .004 = .073	2, 70	2.77	.07
2. BDI	.088 - .046 = .042	2, 66	1.52	.23
3. STAI	.107 - .065 = .042	2, 67	1.58	.21
4. Gender	.078 - .001 = .077	2, 70	2.92	.06
5. Fatigue	.095 - .084 = .011	2, 67	.41	.66
6. Pain	.112 - .102 = .010	2, 68	.38	.68
7. Physical function	.104 - .100 = .004	2, 67	.15	.86
8. Severity	.119 - .113 = .006	2, 68	.23	.79
9. Total (1-7)	.153 - .152 = .001	2, 57	.03	.97

PNN50D

Control Variable	Squared semipartial	DF	F	P
1. Age	.165 - .055 = .110	2, 70	4.61	.01*
2. BDI	.096 - .037 = .059	2, 66	2.15	.12
3. STAI	.170 - .105 = .065	2, 67	2.62	.08
4. Gender	.102 - .001 = .101	2, 70	3.94	.02*
5. Fatigue	.103 - .066 = .037	2, 67	1.38	.26
6. Pain	.204 - .127 = .077	2, 68	3.29	.04*
7. Physical function	.147 - .092 = .055	2, 67	2.16	.12
8. Severity	.186 - .112 = .074	2, 68	3.09	.05*
9. Total (1-7)	.355 - .325 = .030	2, 57	1.33	.27

RMSSDD

Control Variable	Squared semipartial	DF	F	P
1. Age	.209 - .079 = .130	2, 70	5.75	.005*
2. BDI	.096 - .053 = .043	2, 66	1.57	.22
3. STAI	.161 - .116 = .045	2, 67	1.80	.17
4. Gender	.096 - .002 = .094	2, 70	3.64	.03*
5. Fatigue	.082 - .059 = .023	2, 67	.84	.44
6. Pain	.166 - .126 = .040	2, 68	1.63	.20
7. Physical function	.130 - .100 = .030	2, 67	1.15	.32
8. Severity	.149 - .113 = .036	2, 68	1.44	.24
9. Total (1-7)	.355 - .348 = .007	2, 57	.31	.73

SDNND

Control Variable	Squared semipartial	DF	F	P
1. Age	.155 - .020 = .135	2, 70	5.59	.006*
2. BDI	.091 - .016 = .075	2, 66	2.72	.07
3. STAI	.112 - .043 = .069	2, 67	2.60	.08
4. Gender	.090 - .009 = .081	2, 70	3.11	.05*
5. Fatigue	.084 - .067 = .017	2, 67	.62	.54
6. Pain	.134 - .131 = .003	2, 68	.12	.89
7. Physical function	.127 - .125 = .002	2, 67	.08	.92
8. Severity	.127 - .126 = .001	2, 68	.04	.96
9. Total (1-7)	.225 - .223 = .002	2, 57	.07	.93

SDANND

Control Variable	Squared semipartial	DF	F	P
1. Age	.149 - .006 = .143	2, 70	5.88	.004*
2. BDI	.108 - .016 = .092	2, 66	3.40	.04*
3. STAI	.132 - .044 = .088	2, 67	3.40	.04*
4. Gender	.106 - .017 = .089	2, 70	3.48	.04*
5. Fatigue	.105 - .079 = .026	2, 67	.97	.38
6. Pain	.132 - .128 = .004	2, 68	.16	.85
7. Physical function	.126 - .123 = .003	2, 67	.11	.90
8. Severity	.129 - .128 = .001	2, 68	.03	.97
9. Total (1-7)	.204 - .203 = .001	2, 57	.04	.96

LFD5(nu)

Control Variable	Squared semipartial	DF	F	P
1. Age	.121 - .003 = .118	2, 70	4.70	.01*
2. BDI	.096 - .053 = .043	2, 66	1.57	.22
3. STAI	.122 - .079 = .043	2, 67	1.64	.21
4. Gender	.108 - .069 = .039	2, 70	1.53	.22
5. Fatigue	.085 - .043 = .042	2, 67	1.54	.22
6. Pain	.088 - .064 = .024	2, 68	.89	.41
7. Physical function	.083 - .066 = .017	2, 67	.62	.54
8. Severity	.091 - .071 = .020	2, 68	.75	.48
9. Total (1-7)	.178 - .167 = .017	2, 57	.59	.56

HFD5(nu)

Control Variable	Squared semipartial	DF	F	P
1. Age	.152 - .054 = .098	2, 70	4.04	.02*
2. BDI	.094 - .076 = .018	2, 66	.65	.53
3. STAI	.228 - .203 = .025	2, 67	1.08	.34
4. Gender	.115 - .017 = .098	2, 70	3.88	.02*
5. Fatigue	.060 - .044 = .016	2, 67	.57	.57
6. Pain	.128 - .096 = .032	2, 68	1.25	.29
7. Physical function	.069 - .047 = .022	2, 67	.79	.46
8. Severity	.101 - .076 = .025	2, 68	.94	.40
9. Total (1-7)	.400 - .387 = .013	2, 57	.74	.48

PNN625N

Control Variable	Squared semipartial	DF	F	P
1. Age	.206 - .156 = .050	2, 68	2.14	.13
2. BDI	.089 - .020 = .069	2, 65	2.46	.09
3. STAI	.149 - .054 = .095	2, 66	3.68	.03*
4. Gender	.086 - .011 = .075	2, 68	2.79	.07
5. Fatigue	.085 - .009 = .076	2, 66	2.74	.07
6. Pain	.218 - .057 = .161	2, 67	6.89	.002*
7. Physical function	.121 - .016 = .105	2, 66	3.94	.02*
8. Severity	.161 - .029 = .132	2, 67	5.27	.008*
9. Total (1-7)	.369 - .342 = .027	2, 56	1.20	.31

HFN5(nu)

Control Variable	Squared semipartial	DF	F	P
1. Age	.213 - .089 = .124	2, 68	5.36	.007*
2. BDI	.123 - .041 = .082	2, 65	3.04	.05*
3. STAI	.253 - .162 = .091	2, 66	4.02	.03*
4. Gender	.156 - .001 = .003	2, 68	6.24	.003*
5. Fatigue	.122 - .042 = .080	2, 66	3.01	.06
6. Pain	.178 - .083 = .095	2, 67	3.87	.03*
7. Physical function	.139 - .048 = .091	2, 66	3.49	.04*
8. Severity	.155 - .069 = .086	2, 67	3.41	.04*
9. Total (1-7)	.476 - .443 = .033	2, 56	1.76	.18

For virtually all of the variables group differences remained after controlling for Age and Gender. It is interesting to note that the two nighttime measures of HRV (PNN625N and HFN) showed group differences when controlling for all co-variates except Total and even then p=.18. For some variables physical symptoms could not explain group differences and for others psychological symptoms could not. These data seem to indicate that both physical and psychological symptoms play a role in determining group differences.

Comparisons Between HRV During Day and Night

Because of diurnal variations in the autonomic nervous system and the HPA axis, it is of interest to compare the daytime to nighttime values of HRV variables for each group. Table 19 below shows those comparisons for time-domain variables based on paired-t-tests for each variable for each group. Table 20 presents the same comparisons for frequency domain variables

Table 19 – Daytime/Nighttime Comparisons of Time-Domain HRV Variables

Controls		
AVNN	Night>Day	(D=-142.0, p=.000)
PNN50	Night>Day	(D=-4.60, p= .009)
RMSSD	Night>Day	(D=-7.0, p=.036)
SDNN	Day>Night	(D=20.8, p=.000)
SDIDX	No significant Day/Night difference	
SDAN	Day>Night	(D=34.0, p=.000)
PNN625	No significant Day/Night difference.	
FMS/CFS		
AVNN	Night>Day	(D=-121.4, p=.000)
PNN50	Night>Day	(D=-2.9, p= .021)
RMSSD	Night>Day	(D=-4.4, p=.006)
SDNN	Day>Night	(D=14.9, p=.000)
SDIDX	No significant Day/Night difference	
SDAN	Day>Night	(D=22.5, p=.000)
PNN625	No significant Day/Night difference.	
GWI		
AVNN	Night>Day	(D=-136.6, p=.000)
PNN50	Night>Day	(D=--8.2, p= .002)
RMSSD	Night>Day	(D=-10.7, p=.001)
SDNN	No significant Day/Night difference	

SDIDX	Night>Day	(D=-4.9, p=.125)
SDAN	Day>Night	(D=11.4, p=.040)
PNN625	Night>Day	(D=-5.0, p=.014)

Table 20 – Daytime/Nighttime Comparisons of Frequency-Domain HRV Variables

Controls		
TP	No significant Day/Night difference	
VLF	No significant Day/Night difference	
LF	No significant Day/Night difference	
HF	Night>Day	(D=-.3433, p=.002)
LF/HF	Day>Night	(D=1.33, p=.000)
LF(nu)	No significant Day/Night difference	
HF(nu)	Night>Day	(D=-11.4, p=.000)
FMS/CFS		
TP	No significant Day/Night difference	
VLF	Night>Day	(D=0.76, p=.115)
LF	No significant Day/Night difference	
HF	Night>Day	(D=-0.25), p=.042)
LF/HF	Day>Night	(D=1.78, p=.000)
LF(nu)	No significant Day/Night difference	
HF(nu)	Night>Day	(D=-8.12, p=.000)
GWI		
TP	Night>Day	(D=-0.16, p=.078)
VLF	Night>Day	(D=-0.27, p=.012)
LF	No significant Day/Night difference	
HF	Night>Day	(D=-0.55, p=.001)
LF/HF	Day>Night	(D=1.76, p=.000)
LF(nu)	Day>Night	(D=6.34, p=.003)
HF(nu)	Night>Day	(D=-12.50, p=.000)

Generally, there is a trend for all groups to show significant increases in HRV at night as determined by all measures. Additionally, group differences among HRV variables begin to disappear, but FMS/CFS subjects do tend to have lower, though not significantly so, HRV compared to Control. One notable observation is that the LF component in GWI dropped significantly between day and night, whereas it stayed constant in the Control and FMS/CFS groups resulting in do nighttime group differences in LF. In contrast, although there were no differences among Groups in daytime HF, it was significantly lower at night in the FMS/CFS group compared to Control and GWI. This was in spite of significant day/night increases in HF for all three groups. These findings suggest that during the day GWI may have more sympathetically controlled modulation of heart rate during the day, whereas the FMS/CFS group had a smaller restoration of vagal modulation during the night. The catecholamine data (in particular the epinephrine data) may support the finding for the GWI group.

Significance of Autonomic Testing. As previously noted, there is no "all-encompassing" test of autonomic function, and in this protocol, several tests measuring various aspects of autonomic tone and responsiveness were used. Syncope during tilt table testing, when viewed as either a "positive" (i.e. syncope or a significant drop in blood pressure) or "negative" response, was equivocal. We chose this test as a stressor because of the results of early studies by the Hopkins

group demonstrating this to be abnormal in the majority of subjects. Although we found some evidence of differential physiological responses to tilt, the variability inherent in this test has made it of questionable utility. Since our results have not replicated earlier findings, we have abandoned this test in favor of exercise testing in our ongoing protocols.

The heart rate variability analysis, on the other hand, did demonstrate significant differences in both of our CMI (FMS/CFS and GWI) cohorts. These results were largely consistent with those of Martinez-Lavin in subjects with Fibromyalgia [18,19]. The advantage of our data is that all cohorts were performing identical activity (i.e., they were receiving the same set of tests in the GCRC) while these data were being collected, whereas previous studies have not controlled for activity, nor included only un-medicated subjects. The HRV data may bear some relationship to the catecholamine and blood pressures response data from the tilt table test. The GWI group may have had more active sympathetic modulation of heart rate during the day than either the Control or FMS/CFS groups. The observation of a large reduction of sympathetic activity at night in the GWI group may further support this. This could mean that this is a feature that distinguishes the GWI group from the others. The psychological and physical factors that may cause this are not clear. The lack of complete restoration of vagal activity during the night in the FMS/CFS group may represent a distinct phenomenon that separates the GWI from the FMS/CFS groups. We are currently analyzing EKG data collected during tilt see if HRV information during that stressor might provide further insight.

Our next set of studies using this testing will not only examine heart rate variability over long periods of time, but also over short time domains, in response to several stressors. In this manner, we should be able to get a clearer idea of why there is an attenuated release of catecholamines in response to some stimuli such as exercise, and hypoglycemia, but an accentuated sympathetic tone over a 24 hour period. This is not dissimilar to the paradoxical findings with respect to hypothalamic pituitary adrenal function in these conditions, where 24 hour Cortisol has been shown to be low in several studies (although not in this study), whereas individual values at 8AM values are frequently elevated.

Peripheral and visceral pain sensitivity. A dolorimeter evaluation is an accepted measure of peripheral pain sensitivity. Pressure is slowly applied at 22 pre-determined points throughout the body, and the individual is asked when the pressure becomes painful (pain threshold) and when it becomes intolerable (pain tolerance). The units of this measure are kg/cm^2, *and lower values indicate greater tenderness.*

Our findings in FMS/CFS subjects (n = 51) duplicated our pilot data and demonstrated that this group displayed highly significantly increased peripheral pain sensitivity when compared to Controls (n = 49; tender point threshold 1.67 +/- .11 (SE) vs. 2.63 +/- .08, $p<.0001$; control point threshold 2.2 +/- .13 vs. 3.3 +/- .08, $p<.0001$). The dolorimetry values for the entire group of GWI subjects (n = 24) were intermediate between FMS/CFS and Controls, but much closer to the FM/CFS subjects (TP threshold 1.91 +/- .19, Control point threshold 2.51 +/- .22). These results were significantly different from Controls (p = .002, .003 respectively), but not from the FMS/CFS subjects (p = .29, .26 respectively). As expected, the mean number of tender points

was also significantly higher in FMS/CFS subjects, since this is required for the diagnosis (13.1 +/- 4.6 [/18] vs. 4.4 +/- 4.4; $p<.0001$) and the GWI subjects were intermediate and differed from both groups (9.6 +/- 5.8; $p<.01$).

These findings from the entire cohort for each group are particularly striking in view of the fact that greater numbers of GWI subjects were male, and males are known to be less tender than females. When this calculation was repeated in groups matched for gender, the differences were even more striking, again with the GWI subjects being nearly as tender as the FMS/CFS subjects, and significantly more tender than Controls.

Significance. This is the first report to our knowledge demonstrating that GWI subjects display increased peripheral nociception (i.e., a lowered peripheral pain threshold). This obviously represents a highly selected cohort, but supports the notion that similar mechanisms may be operative in GWI, and CMI such as FMS and CFS.

Visceral pain sensitivity. We have used esophageal balloon dilatation as a means of assessing visceral pain sensitivity. A balloon is placed into a standard location in the esophagus, and inflated until the subject senses pain or pressure. Visceral pain sensitivity is known to be decreased in conditions such as irritable bowel syndrome, but is not necessarily abnormal in individuals with illnesses such as CFS and Fibromyalgia (even though these syndromes sometimes coexist).

This value is expressed in cm of air inflated into the balloon. Again, a lower value indicates more sensitivity. Not all subjects were willing to have esophageal testing performed. The amount of air required to cause discomfort in the esophagus was no different in the three groups; GWI (n = 15) 14.2 +/- 0.9 (S.E.) cc, Controls (n=17) 12.0 +/- 1.7, FMS/CFS (n=19) 11.2 +/- .67. There was no relationship between peripheral and visceral pain threshold in the three groups.

Conclusions: This cohort of individuals with GWI displayed significantly increased sensitivity to peripheral pressure stimuli, but not to visceral stimuli. This supports the notion that similar mechanisms may be operative in GWI and FMS/CFS, and also supports previous studies demonstrating that peripheral and visceral nociception are not necessarily related. There are several possible reasons that this group might display peripheral pain sensitivity but not visceral pain sensitivity. The most likely is that the feature of peripheral pain clinically identified them. For example, in IBS, an illness defined in part by visceral hypersensitivity, only approximately 1/3 of patients will also display peripheral hypersensitivity to painful stimuli. Thus, peripheral and visceral hypersensitivity may coexist, but frequently occur in isolation.

Visceral smooth muscle function. Abnormalities in visceral smooth muscle tone have been identified in illnesses such as irritable bowel syndrome, non-cardiac chest pain, etc. that fall within the same spectrum of FMS/CFS/GWI. These abnormalities in smooth muscle tone may contribute to symptoms such as gastroesophageal reflux, diarrhea/constipation, urinary urgency and frequency, etc. Again, it is important to note that in disorders such as IBS, visceral hypersensitivity, and abnormalities in smooth muscle motility, seem to be independent of each

other.

In our study, testing was performed as described using standard manometric techniques and criteria. This was found to be abnormal in 15 of 37 patients with FMS/CFS, with a variety of abnormalities noted including hypotensive lower esophageal sphincter (LES), nonspecific esophageal motor disorder (NSEMD), hypertensive LES. Again, not all subjects consent to this testing, and in the 16 GWI subjects who had this testing performed, ten had abnormal results, with eight of these ten having a hypotensive lower esophageal sphincter. In contrast, these results were only found to be abnormal in 4 of 20 Controls.

Significance. Again, these findings suggest that individuals with GWI display some of the same types of abnormalities in smooth muscle function, in this case of the esophagus, that have been identified in conditions such as irritable bowel syndrome, non-cardiac chest pain, etc.

Overall Conclusions: Many of the data collected in this study support the notion that the same physiologic mechanisms that may be present in conditions such as FMS and CFS, may also be present in ill Gulf War veterans. The most striking finding of this study is that in nearly all of the testing paradigms that identified differences between FMS/CFS subjects and Controls, similar (although quantitatively less prominent) differences were seen between the GWI subjects and Controls. This is true of the studies of the HPA axis, tilt table testing, peripheral nociception, heart rate variability, and esophageal smooth muscle testing. A notable exception to this observation is an increased diastolic blood pressure observed in the GWI group in response to tilt table testing. It is also of note that no testing paradigm identified an abnormality in the FMS/CFS group that was not present (although sometimes to a lesser degree) in the GWI group. Generally, no unique abnormalities were identified in this extensive testing paradigm in the GWI subjects.

The finding of similar physiologic abnormalities in these entities further supports the notion, supported now by numerous population-based studies, that the illnesses that affect Gulf War veterans are not "unique" to the Gulf War. Furthermore, the striking similarities seen in so many different measures, and the absence of any abnormalities seen only in the GWI group, speaks against a unique toxic or infectious cause of illness, at least in the Gulf War veterans in our study.

This is not to say that these studies have "uncovered" the "cause" of chronic multisymptom illnesses such as FMS, CFS, and GWI. These studies help identify potential mechanisms, as well as new avenues of treatment that could be pursued. Also, this multidisciplinary testing paradigm can move one step towards identifying "sub-groups" of individuals with CMI that are based on objective measures (e.g., those with neuroendocrine dysfunction, autonomic dysfunction, primary psychiatric abnormalities) that might respond to different treatments. This would represent a step forward from our current paradigm, wherein subgroups are defined on the basis of arbitrary distinctions (e.g. FMS vs. CFS) that are historical or cultural, rather than real.

There are problems with the present study. First and foremost, our sample sizes were too small, in spite of recruiting larger total numbers of subjects than we anticipated needing based on our initial calculations. A related problem that was brought on by our difficulties with recruitment was that the GWI group was skewed towards being comprised of a high percentage of males, whereas our other groups were composed of primarily females. This, however, reflects the overall gender differences in the populations studied. Another problem was that some of our testing paradigms, turned out to not identify differences between the FMS/CFS and Control groups, thus it is not unexpected that some differences were not found in the GWI group. Finally, our choice of Controls was inadequate to control for factors such as mood, and decondtioning, on some of the variables in question.

These issues have all been addressed in the ongoing studies. Much larger numbers of subjects will be studied, and we are doing so using a different "mind-set", that considers these symptom complexes as occurring over a wide continuum, rather than as discrete conditions. For this reason, we will recruit much larger number of "Control" subjects for study, and include individuals who have moderate, mild, and no symptoms of pain, fatigue, memory complaints, etc. In this manner, we will not just be comparing "hyper-normal Controls" (individuals who never have any pain, fatigue, etc - who may not truly be normal) to subjects at the other end of the continuum (e.g. FMS or CFS). In these individuals with long-established illnesses, there are significant other factors that might influence results on these testing paradigms.

Also, for our ongoing studies we are collaborating with new investigators who can strengthen every facet of our testing paradigms. Foremost amongst this group are the individuals at Walter Reed and UHSUS, who not only bring additional scientific expertise, but also a markedly enhanced ability to recruit symptomatic and asymptomatic Gulf War veterans for our studies.
Our ongoing studies also incorporate several technologies that were not available for this set of studies, such as functional magnetic resonance imaging, novel bioassay techniques, objective assessments of performance, and enhanced cognitive testing. We feel that these technologies can add significantly to our understanding of central nervous system function, where the abnormality is almost certain to be found in these conditions.

Bibliography

1 Institute of Medicine: Health consequences of service during the Persian Gulf War: Recommendations for research and information systems. 1996.

2 Haley R, Kurt T, Hom J: Is there a gulf war syndrome? Searching for syndromes by factor analysis of symptoms. JAMA 1997;277:215-222.

3 Malik M: Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Circulation 1996;93:1043-1065

4 Kaufmann H, Oribe E, Miller M, et al: Hypotension-induced vasopressin release distinguishes between pure autonomic failure and multiple system atrophy with autonomic failure. Neurology 1992;42:590-593.

5 Spitzer RL, Williams JBW, Gibbon M, et al: Structured clinical interview for DSM-III-R.. American Psychiatric Press 1990.

6 Beck AT, Rush AJ, Shaw BFe: Cognitive therapy of depression. Guilford 1979.

7 Ware JE, Serbourne CD: The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Medical Care 1992;30:473-483.

8 Costa PT, McCrae: A six-year longitudinal study of self-reports and spouse ratings on the NEO personality inventory. J Person Soc Psych 1988;54:853-863.

9 Barsky AJ, Barnett MC, Cleary PD: Hypochondriasis and panic disorder. Boundary and overlap. Arch Gen Psychiatry 1994;51:918-925.

11 Melzack R: The short-form McGill pain questionnaire. Pain 1987;30:191-197.

12 Kane R, Kay GG: Computerized assessment in neuropsychology: a review of tests and test batteries. Neuropsychology Review 1992;3:1-117.

13 Crofford LJ. Neuroendocrine abnormalities in fibromyalgia and related disorders. Am J Med Sci 315[6], 359-366. 1998.

14 Bou-Holaigah I, Rowe PC, Kan J, Calkins H. The relationship between neurally mediated hypotension and the chronic fatigue syndrome. JAMA 1995; 274:961-967.

15 Bou-Holaigah I, Calkins H, Flynn JA, Tunin C, Chang HC, Kan JS et al. Provocation of hypotension and pain during upright tilt table testing in adults with fibromyalgia. Clin Exp Rheumatol 1997; 15:239-246.

16 Clauw DJ, Heshmat Y, Groner K, Gondy G, Barbey JT. Tilt table testing in fibromyalgia. Arthritis and Rheumatism 1996; 39(9S):S276.

17 Physiological effects of exhaustive exercise in primary fibromyalgia syndrome (PFS): Is PFS a disorder of neuroendocrine reactivity? Scand J Rheumatol 1992; 21:35-37.

18 Martinez-Lavin M, Hermosillo AG, Rosas M, Soto M-E. Circadian studies of autonomic nervous balance in patients with fibromyalgia: A heart rate variability analysis. Arthritis Rheum 1998; 41(6):354-65.

19 Martinez-Lavin M, Hermosillo AG, Mendoza C, Oritz R, Cajigas JC, Pineda C et al. Spectral analysis of heart rate variability discloses an orthostatic derangement in subjects

with fibromyalgia. Arthritis Rheum 39[9S], 1490. 1996.

RESPONSES TO REVIEWERS OF FINAL REPORT

DAMD #17-96-1-6042
Dysregulation of the Stress Response in
Persian Gulf Syndrome

April, 23 2001

Daniel J. Clauw, M.D.
Principal Investigator

Overview

The reviewers had some general concerns about our approach to statistical analysis of the data. The analytic strategies we employed were selected to provide the clearest comparisons among the study groups while recognizing the limitations imposed by the relative small samples and the confounding effect of some variables (e.g. gender) on our interpretations. It was out of a consideration of these limitations that we did not include all of the possible individual difference variables (e.g. NEO personality scales) in the analyses. Likewise, our decision to form a priori composites in which each component of the composite was given equal weight (for example to assess severity) rather than to undertake multivariate analyses on sets of related variables reflects this consideration. However, in our response here, we have provided post-hoc analyses of individual response variables to demonstrate the meaningfulness of using composite scores. Finally, we elected to use the maximum number of subjects available for many analyses rather than restricting ourselves to only those subjects who had complete data. We appreciate the reviewers concerns about how these decisions may have impacted the analyses and below we address the specific technical issues raised by the reviewers.

Responses to Specific Points

1. The variables in the Tables have been described with their construct names (e.g. depression) and the manner in which they are measured (e.g. Beck Depression Inventory) are indicated in a footnote. Problems with Table numbering have been rectified.

2. The reviewer is correct that this is a standard F-test. Our notation was non-standard because the degrees of freedom were not listed as subscripts. We have corrected this typographical error.

3. The negative values for the control group means reflect how the composite severity index was formed. Because the variables that were used to index overall severity are on different scales we first standard ("z") scored the variables before averaging them. Although we have provided an interpretation of the values in the table in the paragraph immediately above the table,

the reviewer suggested that providing the actual range of values on these standard score composites would aid in their interpretation. The ranges (for the total sample) are

Fatigue	-1.52 to 1.96
Pain	-1.30 to 1.81
Physical Functioning	-1.25 to 1.48
Overall Severity	-1.32 to 1.48.

4. We recognize that we used different labels for Gulf War Veterans with unexplained illnesses. We have now labeled this group consistently throughout as GWVI for Gulf War Veterans' illnesses.

5. The reviewers' concerns about our treatment of outliers are well taken. Outliers can occur for both methodological and substantive reasons. And, regardless of the reasons, they can have an adverse impact on statistical analyses. We elected to drop the outliers from the analysis because their impact on our statistical analyses was to decrease our already low power to the point that we would miss potentially important differences. However, we agree with the reviewer that this was perhaps more expedient than reasonable.

A) Apparent Inconsistency of Studentized Residuals on the Original Page 12.- The apparent inconsistency noted by the reviewer exists because we evaluated outliers at each stage of the analysis. That is, we initially assessed outliers with respect to the total distribution. After eliminating outliers based on studentized residuals for the total distribution we redid the outlier analysis on the new distribution. In this second stage, a score that did not appear as an outlier originally (because of the presence of more extreme scores in that distribution) might appear as an outlier in the new distribution (because that score is now the most extreme). However, the values of the studentized residuals are not comparable across the stages of the outlier analysis because they are based on different distributions. To clarify this we have reordered the outliers and identified that they were determined to be outliers at different stages.

B) Whether to Keep or Eliminate Outliers is Complex.
There are at least four possible analytic strategies.

1) Keep all cases and do the analyses ignoring outliers.
2) Eliminate the outliers before doing the analyses (the strategy we used)
3) Transform the dependent variables to reduce the adverse impact of outliers on the statistical analyses, but keep the outliers in the data set.
4) Use a non-parametric version of the analysis of variance to eliminate the adverse effect of the outliers on the parametric assumptions of the analysis.

Strategies 1, 3 and 4 all assume that the outlying values, although extreme, are valid. That is, they are not the result of measurement problems or methodological or data entry errors. Of these three strategies our preference is to transform the data to reduce the effect of the outlying values on the analysis, but keep those subjects in the data set. However we also did a non-parametric (Kruskal-Wallis) analysis of the group differences on the Cortisol and ACTH analyses.

Ignoring the outliers simply allows these few subjects to have too large an influence on the analysis, either by spuriously increasing the effect size or (as was the case here) inflating the within-group variability.

Thus we have redone the analyses comparing the groups with these two alternative treatments of the outliers.

We used a square root transformation of the cortisol and ACTH variables to reduce the effect of the extreme scores on the analyses. Analyses of the square root transformed values for ACTH revealed no differences among the groups at any of the time points. This result was echoed by the non-parametric Kruskal-Wallis (KW) comparisons. Values of the KW test statistics and associated p values (based on a chi-square distribution with 2 degrees of freedom) were 3.73, $p = .16$, 2.49, $p = .29$, and 3.56, $p = .17$ for Baseline day 1, Baseline day 2, and Midday assessments, respectively.

For cortisol, the analyses based on the square root transformed values and the results based on the non-parametric KW test were again similar to those obtained with the analyses where the outliers were eliminated. For Baseline Day 1 the analysis of variance of the square root transformed values was significant ($F_{2, 88} = 5.37$, $p = .006$) and the KW test statistic was 9.05, $p = .011$). For the baseline day 2 and midday assessments the groups were not found to differ on either the transformed values or on the KW test.

Thus, it does not appear that our removal of the outlying values in the original analysis seriously distorted our results. Again, we recognize that the treatment of outliers with this type of data is not straightforward and we appreciate the reviewer's feedback on this issue. In the revised report we provide this discussion of outliers because, as noted, the issue is too complex to ignore.

6. We very much agree with the reviewer's concerns about statistical power. For the tests of urinary cortisol mentioned on page 13, the effect size based on the data was .14 using Cohen's *f* statistic. This corresponds to a small effect. For reference, a medium effect is .25 and a large effect is .40. The power of this study to detect a small effect is only .11, so the failure to find a difference may be due to low statistical power. On the other hand this study had sufficient power (.87) to detect a large effect, but only very modest power (.25) to detect a medium one. Clearly if small differences in urinary cortisol are of interest this study was not equipped to detect them.

In this regard we note that our interpretation of p values that approach, but do not cross, the .05 threshold as possibly important was made in light of our awareness of the study's power. Thus, the reviewers' later characterization of this interpretation as "more than optimistic" seems a bit harsh.

7. The different degrees of freedom for the various analyses are not desirable. However, if we restricted ourselves to analyzing data only for subjects with complete data on all measures our sample size would have approached zero. Thus we elected to use pairwise rather than

listwise deletion of cases for the analyses. That is, we conducted an analysis on all subjects who were available for that analysis. The problem is compounded because often the data are not missing at random, reflecting differential subject drop out rates or errors in data collection. As a result we could not use a maximum likelihood procedure such as the Expectation Maximization algorithm for estimating missing values. Thus we opted for the use of different subsets of subjects for different analyses.

8. The reviewer raises a number of issues about the correlated covariates in the analyses. It is not clear how a MANOVA would resolve the issue of correlated covariates, as this analysis usually applies to correlated dependent variables. With regard to the dependent variables the values of ACTH are substantially correlated across the three time periods (correlations range from .68 to .74). Cortisol is also correlated across the time periods although somewhat less so (correlations range from .39 to .67). Cortisol and ACTH are generally not significantly correlated in this sample of 34 subjects. However, if we did a MANOVA across the time periods we would have only total 34 subjects available (about 10/group) and this would not be acceptable (see above discussion regarding missing values).

The covariates are substantially correlated as the Table below indicates

	FATIGUE	PAIN	PHYSFUN	SEVERIT	AGE	Depres	GENDER	Anx
FATIGUE	1.000							
PAIN	0.735	1.000						
PHYSFUNCR	0.807	0.829	1.000					
SEVERITY	0.912	0.924	0.947	1.000				
AGE	-0.140	-0.080	-0.147	-0.132	1.000			
Depression	0.622	0.549	0.635	0.648	-0.103	1.000		
GENDER	0.246	0.198	0.167	0.220	-0.124	0.194	1.000	
Anxiety	0.491	0.461	0.520	0.529	-0.157	0.699	0.185	1.000

Note that Severity IS a composite of the Fatigue, Pain, and Functioning variables. Thus, we would not (and do not) include it in the analyses that considered these variables separately. We formed this composite out of appreciation for these high correlations. Our preference is to use the Severity measure rather than its components in the analyses.

Conceptually it makes less sense to form composites of the other covariates, but our use of multiple regression for these analyses adjusts for correlations among the covariates (and predictor variables) in the analyses.

The reviewer makes the intriguing suggestion that severity should be considered as the outcome variable and that ACTH or Cortisol be considered as mediating variables of the group differences in severity. We followed up this suggestion by conducting an analysis to assess the mediating effect of Cortisol measured at baseline day 1 of the relation between group and severity. To establish a mediating effect we need to show 1) that there are group differences in severity, 2) That there are group differences in cortisol, and that when cortisol is added as a covariate in the analysis to assess the group differences in severity, the effect of group decreases. This latter finding implies that some of the group differences can be explained by differences in

cortisol. In these analyses we have used only the severity composite rather than its components.

As shown in Table ? of the report there are substantial differences in the groups on severity with a nearly two standard deviation difference between the gulf war and FM/CFS groups and the control group. Likewise as documented above there are group differences in cortisol assessed at Day 1 baseline. However, when cortisol is added as a covariate to the analysis the relation between group and severity remains significant and the effect size is not materially affected (the standardized regression coefficient for the group contrast that carries the effect changes from .834 to .832). Thus, the intriguing suggestion that cortisol differences mediate the group differences in severity was not supported by these data. Because this is an important hypothesis we have included this discussion in the revised Final Report.

9. The reviewers suggest that we assume that differences among groups of certain measures represent abnormalities. Such an assumption does not form the basis for any conclusions we draw. We recognize that many of our measures, both in certain individuals and in whole groups, are not abnormal per se. Indeed, a major question about CMI is whether the symptom complexes represent the tail end of a distribution including so-called healthy normals, or are they part of a unique distribution representing "abnormal" structure and function. Our operating hypothesis is that persons with CMI do not constitute a unique nosological class, but are part of the overall population distribution for what we might call "normal" functioning. Consequently, we would not expect to see frank clinical abnormalities in CMI subjects. If such existed, we would not be struggling with the underlying mechanisms. We are exploring whether there are consistent differences in baseline function and response of different central nervous system components that could mechanistically explain the presence and severity of symptoms that all humans experience, even in states of good underlying health.

10. The reviewers indicated that it would have been helpful had raw data been displayed. We assume that the reviewers do not mean individual data for each subject. Such a report would amount to hundreds of pages. In the revised report we now include the summary statistics and analysis of variance results for the individual measures that comprise composite measures to help demonstrate post-hoc the utility of composite scores

11. The reviewers suggested that we report sex and age separately for the three groups. We assume that the reason for requesting both medians and means is to assess the extent to which the distributions may be skewed. Table 2 shows the mean and medians of the sex and age variables for each group. Note that although sex is a categorical variable the mean is identical to the proportion of men in each group because sex is scored 0 (women) and 1 (men). The median for sex is not particularly informative.

	Sex			Age		
	Mean	Standard Deviation	Median	Mean	Standard Deviation	Median
Control (N=49)	.12	.33	0	41.8	9.6	45.0
Gulf War (N=25)	.72	.46	1	36.0	8.6	36.0
FM/CS (N = 51)	.06	.24	0	41.1	9.2	44.0

12. The reviewers noted the absence of any mention of a number of other available individual difference measures in the analyses as covariates. Specially the SCID, the SF-36, the NEO, the Barsky, and the COGSCREEN. We tried to be selective in including individual difference variables in the analyses because of the "load" that additional variables would place on our sample size, particularly for the Gulf War group. As our sample size continues to increase we will be able to consider the possible role of additional personality, psychopathology, and cognitive variables in understanding Gulf War Illness. In these analyses we restricted ourselves to continuous and psychometrically sound measures of depression and anxiety because of their psychometric properties and theoretical relevance. In addition there are missing values on most of these individual difference measures, which would again limit our sample size, particularly if we included many of them in the same analysis. However, it should be noted that we used items from the SF-36 in forming the severity composite.

However, we have conducted post-hoc analyses for group difference for the SCID, SF-36, NEO, Barsky, a visual analog scale for pain and fatigue, the McGill Pain Inventory and the Multidimensional Fatigue Inventory. These are given in an appendix to the report and demonstrate an overall consistency with analyses using composite scores.

13. The reviewer asked why cortisol levels were used in developing composites. It is not clear what we said in the report that would imply this because cortisol levels were not used in any of the composites, nor was it considered in anyway in forming the composites.